



Riku Poutanen

ANALYSIS OF ONLINE ADVERTISE- MENT PERFORMANCE USING MARKOV CHAINS

Case Fiksu Ruoka Oy

Tekniikan ja luonnontieteiden tiedekunta
Master's thesis
Maaliskuu 2020

ABSTRACT

Riku Poutanen: Analysis of online advertisement performance using Markov chains
Master's thesis
Tampere University
Industrial Engineering and Management
March 2020

The measurement and performance analysis of online marketing is far from simple as it is usually conducted in multiple channels which results depend on each other. The results of the performance analysis can vary drastically depending on the attribution model used. An online marketing attribution analysis is needed to make better decisions on where to allocate marketing budgets. This thesis aims to provide a framework for more optimal budget allocation by conducting a data-driven attribution model analysis to the case company's dataset and comparing the results with the de-facto last-click attribution model's results. The framework is currently utilized in the case company to improve the online marketing budget allocation and to gain better understanding of the marketing efforts.

The thesis begins with literature review to online marketing, measurement techniques and most used attribution modeling models in the industry. The Markov's attribution model was chosen to the analysis because of its promising results in other research and the ease of implementation with the dataset available. The dataset used in the analysis contains 582 111 user paths collected during 7 months period from the case company's website. The analysis was conducted using R programming language and open source ChannelAttribution package that includes tools for fitting a k-order Markovian model in to a dataset and analyzing the results and the model's reliability. The performance of the attribution model was analyzed using a ROC curve to evaluate the prediction accuracy of the model.

The results of the research indicate the Markov's model gives more reliable results on where to allocate the marketing budget than then last-click attribution model that is widely used in the industry. Overall the objectives of this thesis were achieved, and this study provides a solid framework for marketing managers to analyze their marketing efforts and reallocate their marketing budgets in more optimal way. However, more research is needed to improve the prediction accuracy of the model and to improve the understanding of the effects of budget reallocation.

Keywords: Online Marketing, Attribution, Markov's chain

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

TIIVISTELMÄ

Riku Poutanen: Digitaalisen mainonnan tehon analyysi käyttäen Markovin ketjua

Diplomityö

Tampereen yliopisto

Tuotantotalous

Maaliskuu 2020

Digitaalisen mainonnan tehon analysointi mittaaminen ja analysointi on kaukana yksinkertaisesta tehtävästä koska digitaalinen markkinointi tehdään usein samanaikaisesti monessa kanavassa ja kaikki nämä kanavat vaikuttavat toisiinsa. Analyysin tulokset vaihtelevat merkittävästi käytetystä attribuutiomallista riippuen. Atributioanalyysiä tarvitaan, jotta yritys osaa tehdä parempia päätöksiä siitä mihin kanavaan markkinointibudjettia kannattaa allokoida. Tämän diplomityön tavoitteena on löytää nykyistä vallalla olevaa attribuutiomallia (viimeiseen kosketukseen perustuva malli) parempi tapa allokoida budjettia eri markkinointikanaville. Diplomityössä käytetty attribuutiomalli on otettu käyttöön yrityksessä tukemaan markkinointikanavien välistä budjettien allokointia ja markkinoinnin analysointia.

Työn tekeminen alkoi kirjallisuuskatsauksella digitaaliseen markkinointiin, mittaamistapoihin ja yleisimpiin attribuutiomalleihin. Työhön valikoitui attribuutiomalliksi korkeamman kertaluvun Markovin ketju, jolla on saatu aikaisemmin hyviä tuloksia saman tyyppisissä analyyseissä ja koska se oli sovitettavissa työssä mukana olleen yrityksen keräämään verkkosivujen käyttäjätietoihin. Näytedata sisälsi 582 111 käyttäjäpolkua, jotka oli kerätty 7 kuukauden aikana. Analyysi tehtiin käyttäen R ohjelmointikieltä ja ChannelAttribution vapaan lähdekoodin kirjastoa, joka sisältää työkalut korkeamman kertaluvun Markovin ketjun sovittamiseen näytedataan ja tulosten luotettavuuden analysointiin. Tulosten luotettavuutta ja mallin ennustuskykyä analysoitiin käyttämällä ROC -käyrää.

Työn tulokset osoittavat, että Markovin malli antaa luotettavampia tuloksia markkinointibudjetin allokointiin kuin viimeiseen kosketukseen perustuva attribuutiomalli. Työn tavoitteet saavutettiin kokonaisuudessaan ja tämä työ tarjoaa yrityksen digitaalisesta markkinoinnista vastaaville henkilöille hyvän työkalun digitaalisen markkinoinnin tehon analysointiin ja markkinointibudjetin jakamiseen eri kanavien välillä. Lisätutkimusta vaaditaan kuitenkin, jotta mallin ennustuskykyä voidaan parantaa ja jotta ymmärrys budjetin uudelleenallokoinnin vaikutuksia voidaan ymmärtää paremmin.

Avainsanat: Digitaalinen markkinointi, attribuutiomalli, Markovin ketju

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

PREFACE

I would like to thank everyone that pushed and motivated me during the thesis writing process. At first the journey seemed too long to even start with, but in the end the writing process was considerable reasonable. Special thanks for Tuuli for making sure I started the process and stuck with it even though the thesis wasn't my highest priority and the writing was usually done on Sundays after a long week of work.

I would also like to thank everyone at Fiksu Ruoka Oy so far for the amazing growth story that we have managed to build and making it possible for me to have a such an interesting problem to write a thesis on. Data-driven decision making is one of the most core values for the Fiksu Ruoka Oy team and this thesis is just one piece to the puzzle of building a highly optimized business that decreases food waste in the world and has a positive impact to the world we live in.

Thank you also for my thesis instructor Tommi Mahlamäki for giving me comments throughout the whole writing process, understanding my time schedule challenges and of course for promoting marketing as a field of study in our university.

Lastly, I would like to thank Indecs and its members for the challenges the board year offered me and all the amazing memories during my studies.

Tampere, 21.3.2020

Riku Poutanen

CONTENTS

1. INTRODUCTION	1
1.1 Research scope.....	2
1.2 Structure of the thesis.....	2
2. ONLINE ADVERTISEMENT AND ATTRIBUTION MODELING.....	4
2.1 Digital advertisement tracking.....	8
2.2 Attribution modeling in digital advertisement.....	9
2.2.1 Rule-based attribution.....	11
2.2.2 Data-based attribution.....	14
2.3 Markov's chains in attribution modeling	17
3. DATA COLLECTION AND ANALYSIS METHODS	21
3.1 Data description.....	21
3.2 Data processing and algorithms	23
4. RESULTS	28
4.1 Transition probabilities.....	34
4.2 Removal effects.....	35
4.3 Prediction accuracy and model reliability	35
5. DISCUSSION.....	39
5.1 The value of marketing channels according to the Markov's attribution model	39
5.2 Comparing results of Markov's attribution model to the last-click model ...	40
5.3 Recommendations of budget reallocation	42
5.4 Roadmap for utilizing the attribution model in the case company.....	43
5.5 Limitations of the study.....	44
5.6 Topics for future research.....	45
6. CONCLUSION.....	48
REFERENCES	50

LIST OF SYMBOLS AND ABBREVIATIONS

Attribution model	The principle of assigning credit to advertisement channels
AUC	Area Under the Curve
FPR	False Positive Rate
Classifier	Algorithm that classifies samples to two or more classes
CPC	Cost per click
CPI	Cost per impression
CPM	Cost per thousand impressions
ROC	Receiver Operating Characteristics
TPR	True Positive Rate

1. INTRODUCTION

In the era of online advertisement, the measurement of marketing efforts has become feasible even for small and medium sized companies. However, the results of the user tracking and advertisement platform reports can be misleading and lead to unoptimized results due the lack of understanding the principles in which the results are displayed. Measuring multi-channel marketing is complex even in online channels, because each customer can have multiple touchpoints to the advertisement of the company, and they can visit the website through multiple ads before making the purchase. In digital advertisement the attribution model is the principle of assigning credit to one or more advertisements driving the user to the desirable actions such as making a purchase (Shao & Li 2011). These results are used to make decisions on where to advertise and how much budget should be allocated to each of the channels. If the attribution model is not optimal, the marketing team will make poor decisions and marketing results will not be optimal. Every marketing team wants to allocate their marketing efforts and investments as efficiently as possible. Therefore, the research question of this thesis will be:

“How to make optimal digital marketing channel budgeting allocations?”

The simplest attribution model is to attribute the whole credit to the last touch point in the purchase funnel (Jayawardane et al. 2015). However, this has been proved not to give optimal results (Chandler-Pepelnjak 2009). There are several different attribution modeling techniques and most of them rely on predefined rules such as the last-click model above. However, within the last years the growing amount of data has made it possible to build an attribution model based on historical data of advertisement data and click-stream data from the website. The advantage of this approach is that the attribution model is based on actual behavior of the users in the past and therefore it is argued to give more reliable results when predicting the future actions of the users (Shao & Li 2011). The data-driven attribution model can be used to estimate the amount of conversion each of the marketing channels will bring and it can be used to make sophisticated decisions on how to allocate the marketing budget.

The purpose of the thesis is to research the budgeting allocation principles and how to make the budgeting allocation between different marketing channels more optimal. In this thesis the research question will be examined by conducting a quantitative research

using a data-driven attribution model based on a Markov's chains technique. Originally this method has been used to evaluate the performance of each player in the team on sports such as soccer (Bukiet 1997). Within the last years it has been also adapted in digital marketing and various other applications in data analytics. The Markov's chain model was chosen because it is an emerging attribution modeling technique that has been tested in only a few scientific researches, but it keeps growing its popularity in the digital marketing industry. Also, Markov's model can be conducted based on the data and current tracking setup that the case company is using.

1.1 Research scope

The advantage of the digital marketing channels compared to non-digital channels is the amount of detailed user level clickstream and impression data the channels can provide. For non-digital marketing channels, it is not possible to track which advert triggered the user to visit the website and therefore the impact of these ads is harder to analyze. The scope of the research is limited to the digital marketing channels due the limitations that the traditional non-digital marketing channels have.

The objective of the thesis is to analyze which marketing channels drive customers to convert on the website. The research will not take into account the individual advertisement that the customer saw and handles each of the marketing channels as a homogeneous unit that works with the same efficiency regardless of the specific ad being shown to the user. This is not completely in line with the reality because the performance of each creative varies, but the results can still be used in the decision making of marketing activities such as budget allocation between the channels.

1.2 Structure of the thesis

The thesis will start with introduction to the subject. Chapter 2 will act as a literature review that will introduce previous research on attribution modeling techniques. Chapter 2 will also give an introduction to the marketing channels addressed in the thesis and introduce Markov's chains method.

Chapter 3 will describe the data sources, data structure, data processing and the analysis methods used in the thesis. Chapter 4 will explain the results of the attribution modeling analysis of the case company's website's clickstream data. Chapter 5 will be a discussion of the results. The results of the new attribution model will be compared with the industry standard attribution model (last-click model) that is still being used by most of the digital advertisers. The chapter 5 will also introduce the future research needs, the

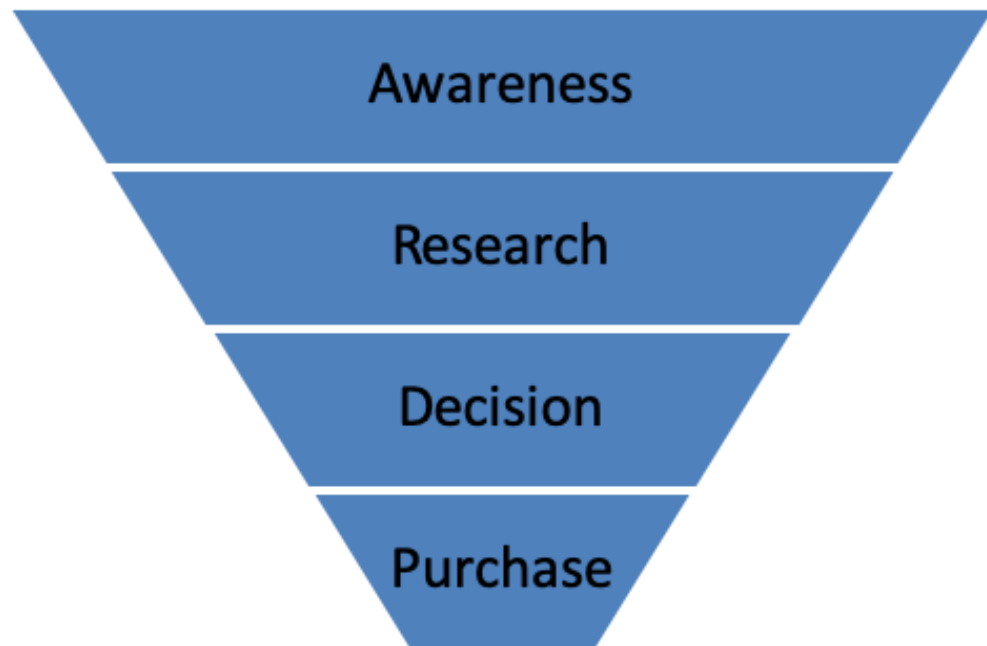
limitations of the research and how should the case company use the information to reallocate the budget to each of the marketing channels. Chapter 6 will conclude the research.

2. ONLINE ADVERTISEMENT AND ATTRIBUTION MODELING

The popularity of internet-based media consumption continues to grow worldwide. As the time spend online continues to grow, so do the advertisement spends allocated to online medias and services. Many online services offer their service to the users for free and monetize the users and their data by showing them ads (Anderl 2014). Advertisers money is the base of the business model and therefore a lot of time and effort is used to develop even better advertisement features and ways to engage users with the ads.

Research conducted PricewaterhouseCoopers (PwC) and published by Interactive Advertising Bureau states that the online advertisement industry revenues in the United States totaled to \$88.0 billion (77,74 billion euros) in 2017 (IAB Internet Advertising Revenue Report 2018). The total growth compared to the year 2016 was 21,4%. Online advertisement surpassed television as the advertisement channel with most revenue in 2016 and keeps on growing while the other channels such as television, radio and print medias have only a little to no growth within the last years. The same kind of growth in digital advertisement revenue has been reported in Europe within the last years according to Interactive Advertising Bureau Europe (European Digital Advertising market has doubled in size in 5 years | IAB Europe 2018).

Advertisement effects on persons behavior have been widely studied, but so far any of the models hasn't been able to fully explain the effects of advertisement in to a person's behavior (Ambler 2000). To properly analyze the behavior of the user affected by the advertisement, marketers have developed various frameworks. The most commonly used one is the funnel approach, often referred as buying funnel, which describes the customer's path to conversion with four steps: awareness, research, decision and purchase (Jansen & Schuster 2011). The conversion means the final outcome of the user behavior that the marketer wishes to accomplish. Depending on the application it can be a purchase, an email capture, a phone call or any other action that is valuable for the marketer. As we move from the start of the funnel towards the conversion point in the funnel the number of customers tends to drop. The same funnel approach is also often referred in website context as a conversion funnel that consists on landing page view, product detail view, add to cart and purchase. The buying funnel is presented in the picture 1.



Picture 1. The buying funnel (Jansen & Schuster 2011)

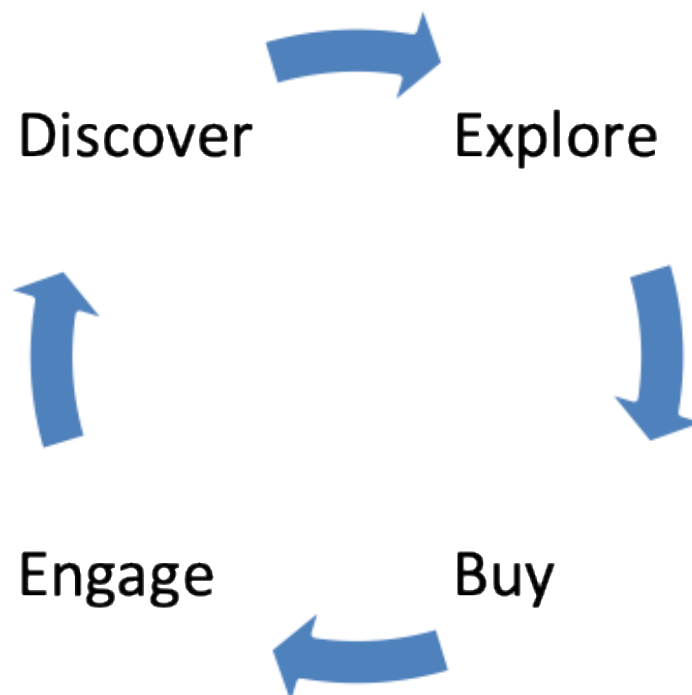
Although the buying funnel approach is widely used among the marketing community, it has been criticized as a too simplified model to a complex problem and by the fact that most customer journeys don't resemble a straight funnel. The research by Jansen and Schuster indicated that the buying funnel model does not represent the actual way that the consumer engages with the brand because the buying process is usually far more complex set of actions and the user can move from funnel stage to another in unpredictable way (Jansen & Schuster 2011). They determined that the funnel approach may not be the most optimal way to describe the online purchase process.

The customer lifecycle model has been proposed as a more modern approach to model the online customer behavior and engagement with the brand and the company (Noble 2010). The customer lifecycle describes the customer journey as a process that has no ending, because it is shaped like a circle. In this model the customer engages with the brand over and over again and moves between the stages called discover, explore, buy and engage.

Table 1 The customer lifecycle phases

Model phase	Description
Discover	The discovery of the brand or product that triggers the new or repeat purchase. For example, seeing an ad or hearing about the product from word of mouth.
Explore	Customer explores the brand and the product and the options in the market. For example, searching information from product pages and reading reviews.
Buy	Customer has made the decision to buy and engages with the purchase process. The process can be interrupted by shortage in stock, bad cart or payment experience.
Engage	After buying the product the customer still engages with the brand in multiple ways: using the product, giving reviews, referring to a friend, warranty, customer support etc.

The customer lifecycle model is argued to be more customer centric model that drives user retention and lifetime value better than the funnel approach. Customer relationship is seen as an ongoing relationship that needs nurturing and every contact point with the brand will affect the lifetime value of the customer. However, there only a little academical proof that this model a better model to describe the online customer behavior than the funnel model. The customer lifecycle model is presented in the picture 2.

**Picture 2.** The customer lifecycle model (Noble 2010) on the right

Modeling the user behavior with a simple model can be a very demanding task also because of the various touchpoints that the user usually has to the brand before the

purchase decision. In today's digital world the amount of advertisement one person sees is overwhelming. Although there are some techniques that allow the marketer to track a customer's interaction with the company's ads in impression level (called impression tracking), the techniques usually can't reliably count every impression the user gets due the difficulty in cross-device tracking. Also, the impression tracking only counts for online advertisement and the offline channels can't be measured in such a detailed way.

Studies of ad impressions and the user's behavior with increasing number of impressions from the same ad are usually conducted in single channels because of this measurement problem. A study conducted by Arrete et al. (Arrate et al. 2018) with a very large dataset of Facebook advertisement data showed that an average Facebook user is exposed to 70 ads from 22 advertisers every week. The study indicated that the probability of a user interacting with the ad is largest on the first impression and decreases as the ad is shown more. Based on this information, if advertiser wants to maximize the effect of the advertisement they should optimize towards reach. However, the study also indicated that the overall probability of user engagement keeps growing logarithmically with number of impressions and therefore larger frequency can be justified. Showing the ad multiple times to the same user with decreasing probability to engage with the ad increases the cost of advertisement. The same wear out effect, also known as ad fatigue (Abrams & Vee 2007), was also discovered by Chatterjee et al. in their study of website clickstream data (Chatterjee et al. 2003).

There are multiple ways to price the online advertisement. The most popular pricing models are either tied to the number of impressions or the clicks to the advertisement. These models are called cost-per-impression (CPI), cost-per-mille (cost per thousand impressions, CPM), cost-per-click (CPC). Most of the online advertisement platforms nowadays work with an auction-based system that enables the marketer to bid a certain amount of money for given metric, for example a click. The advertisement platform combines the bid to the information about the ad's former performance (ad relevancy) and the user's former behavior and interests to estimate the probability of the users clicking the ad and then chooses the ad with most points to be shown to the customer. This system is argued to improve the relevancy of the ads to the user and therefore improve the user experience and also guarantee higher click through rates. The auction system makes sure, that running non-interesting ads will cost the company more and therefore the quality of the ads and the targeting will be better from the customer point of view. (Facebook 2019)

As the prices of ads depend either on the impressions or the clicks and not the actual end conversions such as a purchase, it is the marketer's responsibility to calculate a

correct bid amount. Even though some advertisement platforms charge based on the impression or click they also nowadays offer optimization possibilities that automatically optimize the bid amount in to a specific cost goal for deeper funnel action such as a purchase. This makes it easy for the advertiser to make sure they are not spending significantly more than they are willing to pay for one conversion.

The pricing models of the advertisement platforms have raised concerns around so called click frauds which mean driving up the costs for certain advertiser to exhaust their budget or to increase the advertisement platform revenues. However, the study of Wilbur and Zhu proved that it is not beneficial for the advertising platform to allow click fraud as the advertiser will lower their bid accordingly to the actual results they get from the advertisement (Wilbur & Zhu 2009).

2.1 Digital advertisement tracking

Online advertisements user tracking is usually performed using cookies, which are small files stored in user's browser. The cookie is often referred as a pixel, because it is usually a single-pixel image that is fetched from the server to the user's browser, but not actually displayed in the website. The cookie can be used to identify a user that visits the same site multiple times with the same browser. After placing the unique cookie information to the user's browser, the same identifier is stored in the website's server or website analytics platform such as Google Analytics. The advertisers can use this stored cookie information to conduct analysis of user behavior in the website and retarget the user with relevant ads of the products the user browsed. (Google Developers: Cookies and User Identification 2018)

The problem with the cookies is that they are browser specific. This means that if the same user visits the same website using a different browser or device, the user will be identified as a new user. Nowadays the problem can be solved by if the user can be identified as the same user in the website for example asking the user to log in and then passing a unique user identifier called user id with the cookie. In that case the website analytics platform can identify that the two different cookies in different browsers are related and belong to the same user. (Google Developers: Cookies and User Identification 2018)

In 2018 Google published a new feature that enables advertisers and website owners to exploit the huge amount of data Google has on the users that have been logged in to Google accounts Such as Gmail, YouTube, Google+ (Google Analytics Help 2019) . These accounts enable Google to follow users across multiple devices and Google

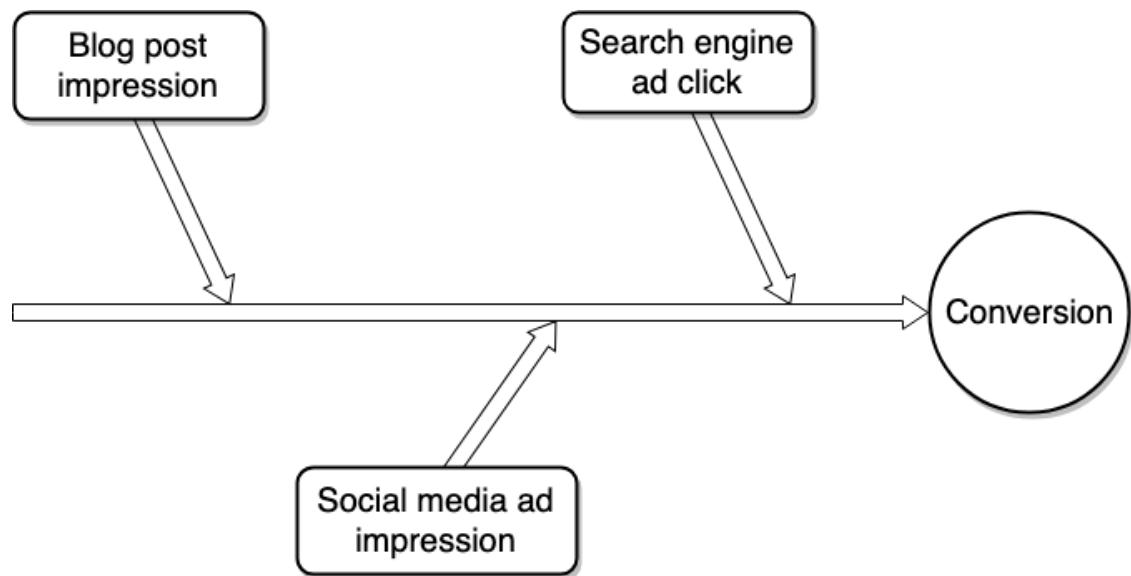
started to share more of these insights with Google Analytics users. This feature provides website owners more reliable data on which sessions are actually unique users and which sessions just the same users browsing with different device.

2.2 Attribution modeling in digital advertisement

Each of the digital marketing channels can contribute to the purchase process in a different way; some of the channels build awareness of the product and brand, some offer more information about the product and some guide the customer to the webstore to make the purchase. As there are endless amount of different marketing channels and not all of them contribute in the conversion event in a same way, the advertiser needs to know which channels increase the probability of the conversion more than the others. By knowing this the advertiser is able to optimize the marketing efforts and allocate more budget and resources to the better performing channels. Allocating the credit for each of the advertisement touch points before the conversion can be done in very many ways. In the literature this problem is called the multi-touch attribution problem (The Interactive Advertising Bureau 2016) .

The Interactive Advertising Bureau defines the attribution as: “Attribution is the process of identifying a set of user actions (“events”) across screens and touch points that contribute in some manner to a desired outcome, and then assigning value to each of these events” (The Interactive Advertising Bureau 2016). The touch point can be considered as an impression or a click depending on measurement technique being used.

The multi-touch attribution problem can be illustrated with the following example. First, the customer reads about the new brand and it’s offering in influencer blog post made in paid collaboration with the brand. Second, the same customer sees an ad in social media and decides to look up more information about the brand in a search engine. The customer ends up clicking the first search result which is a paid ad and ends up in the company’s website. The customer makes the purchase and the desired outcome (conversion) is achieved. The same example of the conversion path is presented in the picture 3.



Picture 3. Example of a customer journey

According to Shao and Li, the multi-touch attribution is one of the most important problems in digital advertisement especially if there are multiple different types of channels involved in the marketing mix (Shao & Li 2011). Some commercial multi-touch attribution solutions exist, but the large variety in the techniques used suggests a present lack of standardization in the industry (Dalessandro et al. 2012). The calculated performance of each marketing channel can vary drastically depending on the attribution model chosen. Therefore, the marketing efforts and budgets can be directed to wrong channels if the attribution model does not model the actual behavior of the users and credits the wrong channels for the conversion. The companies that don't succeed in the attribution modeling waste money and resources on the advertisement that creates no actual results. Even though the attribution modeling is a key for optimized marketing, and it is in great interest by marketing managers, there is no standardized methodology to model the cross-channel effects in the industry (Shao & Li 2011; Anderl et al. 2014).

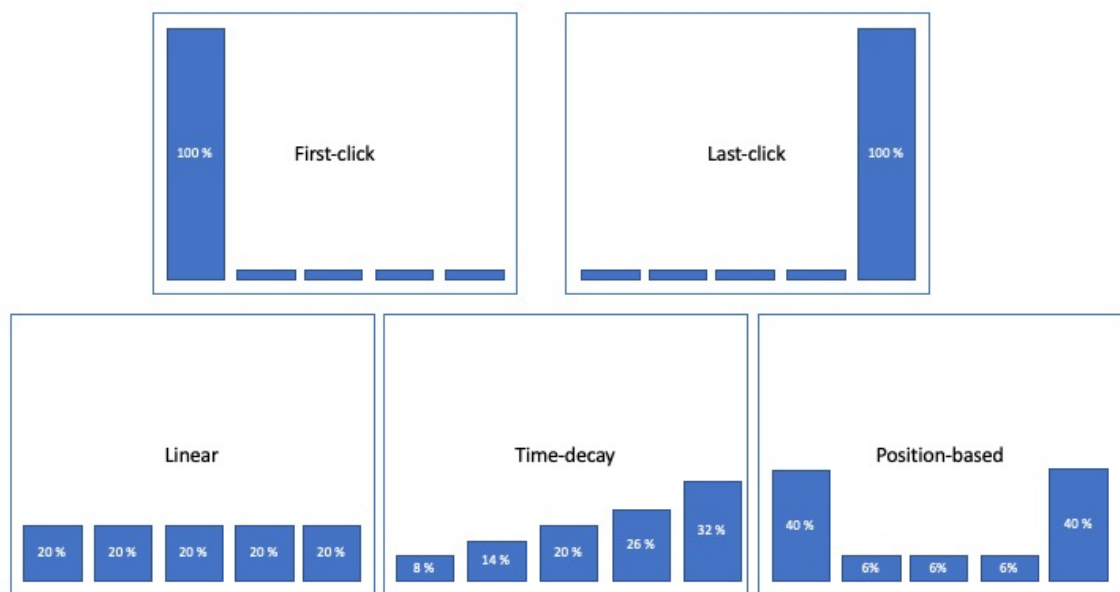
All the of the digital advertisement attribution models consider only digital channels, which may flaw the results if the company is also using a lot of offline advertisement. So far developing an attribution model that would also consider offline channels would be nearly impossible because the offline channels can't be measured on user level. Due the difficulties in the measurement of offline channels, the literature tends to focus only on measuring the online channels.

Academical research on the attribution modeling has emerged within the last years, but so far, the sophisticated attribution models have not found wide acceptance in practice. To be widely accepted within the industry, the attribution modeling methodology must be objective when evaluating the performance of the channels, be able to predict the future

conversion events, be consistent in delivering the results, be transparent and easily explainable to all stakeholders, be easy to adapt to new information and data, be flexible to configure to company-specific use cases and be efficient in computing the results. (Anderl et al. 2014)

2.2.1 Rule-based attribution

Rule based attribution models, also known as heuristic models, rely on a predefined set of rules on how to divide the credit between the advertisement channels on the customer journey. The most popular rule-based attribution models are presented in the picture 4. The picture presents a five-step customer journey, but the same rules can be applied to customer journeys with any length.



Picture 4. Rule-based attribution models

The rule-based attribution models are popular due the ease of implementation and understandability, but the rules tend to simplify the actual behavior of the customers to some intuition-based model that is not derived using a real data collected from the users. (Jayawardane et al. 2015). This chapter introduces the most popular rule-based attribution models.

Last-click model

The most widely used methodology to attribute conversion is the *last-click* model, which assigns all the credit from the conversion only to the last touch point in the customer journey (The Interactive Advertising Bureau 2016)(Jayawardane et al. 2015). The popularity of the model can be explained with the ease of implementation and because the

model is very easy to understand. Because of the popularity of the model, it is often considered as the base model for comparison in the digital marketing literature (Jaya-wardane et al. 2015). Most major digital advertising measurement platforms use this as the default model to track conversions. In some cases, where the customer journey is quite short, the use of the last-click model can be justified. However, as the model only takes in to account the last touch point in the customer journey, it ignores large amount of information and therefore tries to oversimplify the customer's behavior. It has been widely discussed in the literature that the last-click model tends to favor some of the channels which usually locate in the last steps if the customer journey and tends to under valuate the channels that act in the discovery phase of the customer journey (Chandler-Pepelnjak 2009). For example, Kireyev et al points out that banner advertisement drives large amount of the search engine queries (Kireyev et al. 2013). As the customers click on the paid advertisement in the search engine and end up buying, the credit would be assigned to the search engine although the display ad played considerable role in the customer journey and the buying decision. This leads to unoptimized budget allocation and marketing decisions.

First-click model

The *first-click* model is very similar to last-click model, but instead of the last advertisement in the customer journey the credit is assigned to the first advertisement channel the user clicked or saw (The Interactive Advertising Bureau 2016). This model can be used to see which channels are driving brand awareness and act in the beginning of the customer journey. However, it is not as good model to identify channels that actually affect most on the customer before the purchase decisions. Also, defining the correct time interval to measure can be tricky as choosing a too short interval tends to leave the actual first channel from consideration and on the other hand choosing a very long interval can be misleading as the customer may not even remember seeing or clicking the ad anymore. Also, the first-click model results can be flawed due measurement difficulties with multi device customer journeys and users removing tracking cookies or the lifetime of cookies ending (The Interactive Advertising Bureau 2016). The first-click model can also lead to unrealistic results if the company is using a lot of offline advertisement, which is not considered in the attribution modeling and therefore the credit is assigned to the first online touch point.

Linear model

As argued, the one-touch models (first-click and last-click) over simplify the buying process and can therefore lead to unoptimized budget allocation and marketing effort re-sourcing. For these reasons several multi-touch models have been proposed. The simplest one is called *linear-model*, which credits all of the touch points in the customer journey with equal credit (The Interactive Advertising Bureau 2016). Also, the linear-model tends to over simplify the buying process as it credits each of the channels with equal credit although some channels may play more significant role in the purchase decision. However, at least the model considers all the available data and can be therefore a good option to the last-click or first-click models.

Time-decay model

It has been proven by Chatterjee et al that the ads have a wear out effect meaning that the effects of the advertisement decay with time (Chatterjee et al. 2003). For this reason, a *time-decay* model has been proposed to give more credit the latest impressions and click in the customer journey (Jayawardane et al. 2015). There are multiple different weightings and time periods for the model, but usually the touch point on the conversion date receives twice the credit than the touch point that occurred 7 days prior the conversion. The results of the model are flawed for very long and very short customer journeys where the weightings are not distributed fairly between the channels that played a role in the buying decisions. Therefore, when using the time-decay model the time to conversion should be analyzed and the attribution model should be adjusted based on the normal customer behaviors in that industry. For example, in the retail industry the decision-making time is usually quite short but for luxury products the decision making can take much longer and the attribution model should be adjusted accordingly.

The position-based model

As the last-click and first-click models both have their own strengths in emphasizing different kind of channels in the customer journey (the buy and the discovery phases) a *position-based* model, also known as U-shaped model, has been proposed to combine the strengths of the two models (Google Attribution Playbook 2012). The idea behind the model is simple: to give most of the credit (for example 40%) to the first touch-point that triggered the interest and 40% to the last point that resulted in to a conversion. The rest of the credit (20%) is divided equally to the channels that acted as a reminder in the customer journey. Although this model takes in to account various channels it has similar shortcomings as the other rule-based models. The time window taken in to account may not include all the touch points in the customer journey, or it may over emphasize touch

points that occurred too far ago and didn't actually have an effect to the purchase decisions. As all the rule-based models, the position-based model does not take in to account the real behavior data and therefore the results may not model the actual behavior of the customers. For these reasons data-driven attribution models have been proposed.

2.2.2 Data-based attribution

Data-based attribution models, also known as algorithmic or data-driven models, represent a more sophisticated attribution modeling techniques that exploit real clickstream and impression data to build the attribution model (Jayawardane et al. 2015). Unlike rule-based attribution models that rely on set of predefined rules to model the user behavior, the data-driven attribution models offer a probabilistic approach to the problem and therefore offer a less biased view than the models that include build in assumptions (Rentola 2014). A probability of the user converting after interaction with each of the marketing channel is calculated and this probability is used in consideration when dividing the credit to the channels. The channels with larger probability of converting after an interaction will get more credit for each interaction than the ones with smaller probability. In optimal case, the data used to build the model should be collected from the company under evaluation so that it represents real customer behavior as accurately as possible. Data-driven attribution modeling also allows to build different attribution models for different customer segments (Jayawardane et al. 2015) .

It comes without saying, that the models need certain amount of data to reliable estimate user behavior and the reliability increases as the amount of data increases. Therefore, the data-driven attribution models are not an option for very small advertisers. For example, Google requires 600 conversions and 15 000 click in Google Search engine to build their data-driven attribution model for search advertising campaigns (Google Ads Help). These requirements are easily matched by the case company's dataset and the attribution modeling is therefore feasible with a data-based attribution model.

Logistic regression model

Logistic regression is one of the earliest data-driven attribution techniques that was first introduced to marketing community by Chatterjee et al. in their article "Modeling the Clickstream: Implications for Web-Based Advertising Efforts". In the article Chatterjee et al. studied the effects of repeated banner exposure to customer click proneness based on clickstream data of banners in one website. Later the same technique has been applied to solve multi-touch attribution problem by many academics such as Shao and Li (2011) and Rentola (2014).

A binary logistic regression is often used in classification problems to classify the observations in two distinct classes based on the information available. In multi-touch attribution problems these classes are usually converting and non-converting customers. A classification problem of classifying an observation to two or more classes can be solved with numerous different techniques including vector machines, random-forest and neural networks. The problem with these models is often that they build a black box type solution that is not easily understandable for the marketing decision maker and therefore can't be easily applied to marketing optimization decisions. For this reason, Shao and Li decided to use a simple binary logistic regression model but improve the reliability of the results with technique called bagging, which increases the amount of training data points and therefore prevents overfitting and increases prediction accuracy (Shao & Li 2011). The results in prediction accuracy were very similar than with normal binary logistic regression, but the variance of the results was much smaller, which is desirable for attribution modeling.

Shapley value model

Dalessandro et al. proposed a causal framework attribution method to solve two limitations that arise with logistic regression models; the intuitively hard to interpret coefficients and the negative coefficients that may arise due to the channel collinearity (Dalessandro et al. 2012). Dalessandro et al. state to be able to produce a fully unbiased estimation of the causal effects of advertisement the data should meet 3 very strict assumptions, which are: the ad treatment happens before the outcome, any attribute that may affect the ad treatment and conversion outcome is observed and accounted for and that every user has a some non-zero probability of receiving an ad treatment. They also conclude that in a multi-channel marketing campaign it is highly likely that the second and third assumption will be violated and therefore proposed a simplified approximation technique based on game theoretic framework. This multi-touch attribution model is called the Shapley Value regression.

The Shapley value is a method developed by Nobel prize winner and game theorist Lloyd Shapley (Shapley 1953). The Shapley value method was originally developed to model each player's contribution in a cooperative game, but it has been later applied in various other fields such as advertising. The Shapley value treats each of the advertisement channels as a player in the game and assumes that they all play together to influence the customer to convert (Zhao et al. 2018). The method takes the weighted average of the channel's marginal contribution over all possible combinations of the channels (Zhao et al. 2018). The value therefore models the channel's contribution to the total conversions alone and together with other channels.

Dalessandro et al. also tested the model on various real advertisement datasets and compared the results with last touch model. As the real truth on which channel affected most on the conversion cannot be indicated, only differences to other attribution models can be evaluated by comparing the results of two or more models. The results of the comparison depend on the advertisement channels used and the business environment in evaluation so the differences between two models can be drastically different for different companies and data sets. For this reason, Dalessandro et al. also ran a simulation test where they could set the parameters driving ad propensity likelihood (likelihood of the ad being shown), simulated conversion rate for the channel and the last touch propensity (the likelihood of the channels being the last one in the advertisement funnel). The simulation indicated that the last-touch model was mostly driven by the last touch propensity and the multi-touch model was driven by the simulated conversion rate. The simulation proved that the multi-touch model attributes most reward to the channels that actually drive the conversion likelihood and therefore it is more in line with advertisers' objectives than the last-touch model. After Dalessandro et al. introduced the Shapley value attribution model in advertisement context, it has been used by many academics (Sebastian Cano-Berlangaa et al. 2017) (Zhao et al. 2018) and in commercial solutions. For example, Google uses the Shapley value method in their data driven attribution model (Data-driven attribution methodology 2019).

Although the Shapley value method has achieved some industry acceptance within the academics and commercial solutions, it has some shortcomings. The model expects that the sequence of the channels does not affect the results. This is not very in line with actual customer journey, as it can be argued that the results can vary a lot based on which order the channels contributed in the customer journey. Also, the Shapley value calculation is a computation intensive task (Zhao et al. 2018) because the Shapley value must be calculated 2^n times, where n is the amount of advertisement channels and therefore calculating the Shapley value for 15 or more channels is nearly infeasible and at least a very timely manner. This also means, that the Shapley value is not a good model if the effects of advertisement would be analyzed between different campaigns instead of channels as there are usually many campaigns running in one channel and therefore the amount of Shapley values would grow too large.

Markov's chains model

Another data-driven model has emerged within the last years and it has grown popularity especially within the data science community. The attribution method is called Markov's chains. Markov's chains patches some of the argued shortcomings of Shapley value as it takes in to account the sequence in which the channels appeared, and the calculation

is not as computation intensive so it can be even calculated on campaign level. This thesis uses the Markov's chains methodology to model the effects of marketing channels (channel attribution problem).

2.3 Markov's chains in attribution modeling

Markov's chains were first applied to digital marketing attribution modeling by Archak et al. in context of search engine advertisement (Archak et al. Apr 26, 2010). Later Anderl et al. proposed first and higher order Markov's chains to be applied in channel attribution problem (Anderl et al. 2014)(Anderl et al. 2016). According to Anderl et al. the Markov's chain methodology meets the most important requirements for attribution model: objectivity, predictive accuracy, versatility, interpretability, robustness and algorithmic efficiency.

In order to apply Markov's chains in a channel attribution problem we need to present each customer journey as a sequence of touch points to the marketing channels (c_1, c_2, \dots, c_n) that the customer has encountered before the conversion. Markov's chain treats these marketing channels as states (s_1, s, \dots, s_n) in the customer journey

$$S = \{s_1, s_2, \dots, s_n\} \quad (1)$$

and combines the states to the transitioning matrix W with transitioning probabilities between each of the states

$$W = P(X_t = s_j | X_{t-1} = s_i), 0 \leq w_{ij} \leq 1, \sum_{j=1}^N w_{ij} = 1 \forall i \quad (2)$$

where W is the transitioning probability to the next state s_j given the current state s_i . The transitioning probability w_{ij} is always between 0 to 1 and the sum of all transitioning probabilities from on state is 1. A Markov's chain is a sequence of states that represent individual customer journey. A Markov's graph is a representation of all the states and the transitioning probabilities $M = \{S, W\}$. (Anderl et al. 2014)

As a simple example, if the marketing mix of a company would include three channels C1, C2 and C3, the states included in the Markov's chain would be S1, S2 and S3. For modeling purposes, we would need to add 3 more steps (START, CONVERSION and NULL) to be able to fully present the customer journey and calculate the conversion probabilities between every channel. The START state represents the starting of the customer journey, the CONVERSION state the successful conversion and the NULL state the end of customer journey that hasn't ended in successful conversion within the observation time window. The transitioning probability w_{ij} means the probability that the contact with state i is followed by contact with state j . In the simple example our company

would have a dataset of three customer journeys as presented in table 2. Picture 5 shows an example Markov's graph based with the transitioning probabilities calculated based on the three customer journeys presented in the table 2.

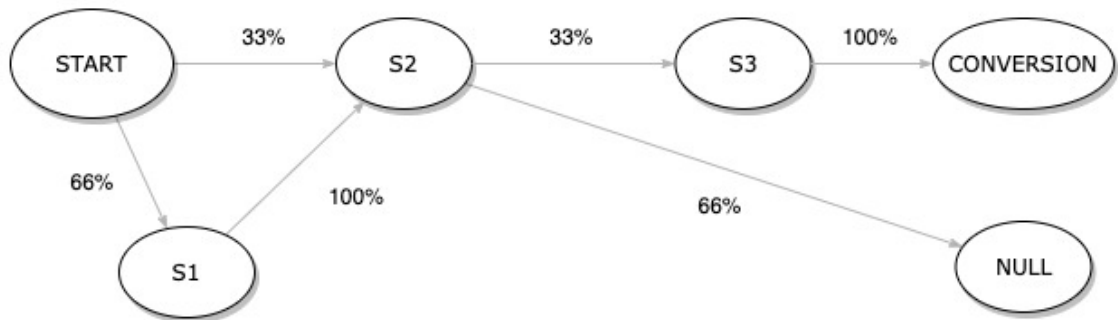
Table 2 Example customer journeys

Journey name	Journey states
Journey 1	START, S1, S2, S3, CONVERSION
Journey 2	START, S2, NULL
Journey 3	START, S1, S2, NULL

The order of the Markov's chain defines how many states before the current state is considered when calculating the probability (Anderl et al. 2014). First order Markov's chain takes into account only the current state and the probability of going anywhere from that state. Second order Markov chain looks back one state, so it takes in to account current state and one state before the current state. Third order Markov chain looks back two states and so on. The transition probability of a k-order Markov's chain is calculated as follows:

$$w_{ij} = P(X_t = s_t | X_{t-1} = s_{t-1}, X_{t-2} = s_{t-2}, \dots, X_{t-k} = s_{t-k}). \quad (3)$$

As the order of the Markov's chain increases, the number of independent parameters and complexity of the model increases exponentially and therefore the modeling task becomes more computing intensive and the risk of overfitting the model increases. Although higher order Markov's chains tend to model the customer journeys more accurately, at some point the model becomes too complex for real world datasets that are usually limited by size (Anderl et al. 2016).



Picture 5. Example of a Markov's graph with transitioning probabilities

To be able to use Markov's chains technique for attribution modeling Anderl et al. (Anderl et al. 2014) propose a removal effect analysis to be used to determine the change in the

probability of reaching from START state to the CONVERSION state if one of the marketing channels is removed. This change in the overall probability models the effect of each marketing channel in the company's marketing mix. For example, we can calculate the removal effect for the channel S1 in the example customer journeys in the table two by calculating the probability of reaching the conversion when we remove the channel S1 from the model:

$$\begin{aligned}
 P(\text{conversion after removing } S1) &= \\
 P(S2 \rightarrow S3 \rightarrow \text{conversion}) &= \\
 0,33 * 0,33 * 1 &= 0,11
 \end{aligned} \tag{3}$$

This means that we can convert 11% of all conversions if we remove the channel S1 from the marketing mix. With the S1 channel intact we can convert 33% of all customer journeys and therefore the removal effect of channel S1 is $0,11/0,33 = 0,33$. This means we would lose 33% of all conversions if we remove the channel S1.

From the Markov's graph in the picture 5 it is easy to see that all the customer journeys that lead to conversion use the channels S2 and S3 and therefore their removal effect is 1, meaning that if we remove the channels, we would lose all the conversions. Now that we know all the removal effect, we can calculate how much credit should each of the channels get from the total amount of conversions based on channels relative removal effect compared to the removal effects of other channels. For example, the S1 channel's attribution coefficient is calculated $0,33 / (1+1+0,33) = 0,14$. This implies that we should attribute 14% of all conversions to the channel S1. It is good to note that this is a highly simplified example of customer journeys as there are usually hundreds or even thousands different customer journeys especially if the company uses high number of marketing channels to drive traffic to the website.

Calculating the first order Markov's chain results is a fairly simple task, but the complexity increases as we use higher order Markov's chains. Higher order Markov's models allow us to calculate the removal effect on states representing channel sequences and therefore take the order of the touchpoints to marketing channels in to account. In such cases the effect of a single marketing channel is calculated as a mean of the removal effects of all the states having that specific marketing channel as the last channel in the sequence.

Markov's chain was selected as the analysis method for this thesis due the good results in prediction accuracy (Anderl et al. 2016)(Alblas 2018), the algorithmic efficiency that allows continuous recalculation of the model and the ease of implementation with widely

used data collection service (Google Analytics) data and the ChannelAttribution R programming language package.

3. DATA COLLECTION AND ANALYSIS METHODS

This thesis examines data set from a company that operates in Finnish fast-moving consumer goods industry. The company sells products only through an online store and therefore has no physical stores. The company is fairly young, and the marketing efforts focus mainly on attracting new customers mainly through digital channels such as social media, search engines, email and display networks. This thesis is limited to studying online marketing channels due the limitation in measurement possibilities of offline channels. Although the company has conducted some offline marketing campaigns during the data collection time period, the investment in offline channels has been rather low compared to the investments in online channels and therefore the effects of the offline channels should not interfere the results.

The purpose of this thesis is to get as objective view of the case company's marketing efforts and analyze the causality between marketing efforts and purchases (conversions). Therefore, the research will be conducted as a quantitative research by analyzing the advertisement and click-stream data of the company. The research is based on positivism philosophy and every result and conclusion must be based on data and objective analysis. Due the data intensive nature of the research, the positivism the most suitable research philosophy for this thesis.

A k-order Markovian model is fitted to the dataset to find statistical correlations between customer paths and conversions. The reliability of the analysis is analyzed using receiver operating curve (ROC) methodology.

3.1 Data description

The data used in the analysis was collected between November 2018 and May 2019 using a free version of Google Analytics software (Google 2019) which is widely used in the digital marketing industry in firms all size to collect user behavior data in websites. The dataset contains individual level clickstream data of all the traffic sources the user arrived to the website from. The same user can arrive to the website multiple times within the time period of observation (lookback window) of 30 days and the sequence of sources represents the customer journey of the user. The dataset contains all customer journeys that ended up as a conversion, but also the ones that didn't end up in a conversion.

Google Analytics offers a standard report of the conversion paths in the website. This report usually only contains the so-called ecommerce conversions that ended up in transaction, but it also offers a possibility to analyze custom goals as conversions. The case company has a website visit defined as a custom goal which enables us to also analyze the paths that didn't end up in transaction. Additionally, a custom conversion segment was used to separate the two groups of users with conversion paths ending up in transaction (later called converting users) and not in transaction (later called non-converting users).

Table 3 presents two example rows of different user paths and the amount of conversions and conversion value (revenue) they have resulted in. The table also contains the number of non-conversion sessions for the user path. The dataset consists on total of 582 111 user paths. Table 4 presents information about the datasets size.

Table 3 The user paths dataset example

User path	Conversions	Conversion value	Non-conversion sessions
A > B > C	2	98,05 €	24
B > C > A	1	23,10 €	28

Table 4 Dataset metrics

Description	Dataset
Total user paths	582111
Unique paths	78224
Conversions	28619
Non-conversions	553492

The data is collected using a default Google Analytics setup which means that the limitations of cookie-based tracking apply. This means that a same user could have used different devices to enter the website at some point and the sessions are separate rows in the dataset. Any sessions prior to the lookback window of 30 days before the conversion will not be accounted for in the dataset. The traffic sources in the dataset is grouped to relevant entities which are listed in table 5.

Table 5 The marketing channels used in the analysis

Entity name	Description
Direct	Direct traffic to the website
Email	Traffic from email newsletters
Generic paid search	Paid search engine traffic without search queries with brand keywords
Branded paid search	Paid search engine traffic with brand keywords
Organic search	Organic (non-paid) search engine traffic
Paid social	Traffic from paid ads in Facebook and Instagram
Organic social	Non-paid traffic from Facebook, Instagram, Youtube and other social medias
Referral	Traffic from other websites not listed above
Snapchat	Traffic from Snapchat ads
Criteo	Traffic from Criteo display network
YouTube	Traffic from YouTube advertisement
Affiliates	Traffic from affiliate partners and influencer posts
Display	Traffic from display ads in Google display network
SMS	Traffic from text message advertisement
Other	Other traffic that does not belong to any of the other channels above

The entities are defined for the case company by taking in consideration the marketing investments and resources so that the company can make marketing budget and time allocation decisions based on this attribution analysis in the future.

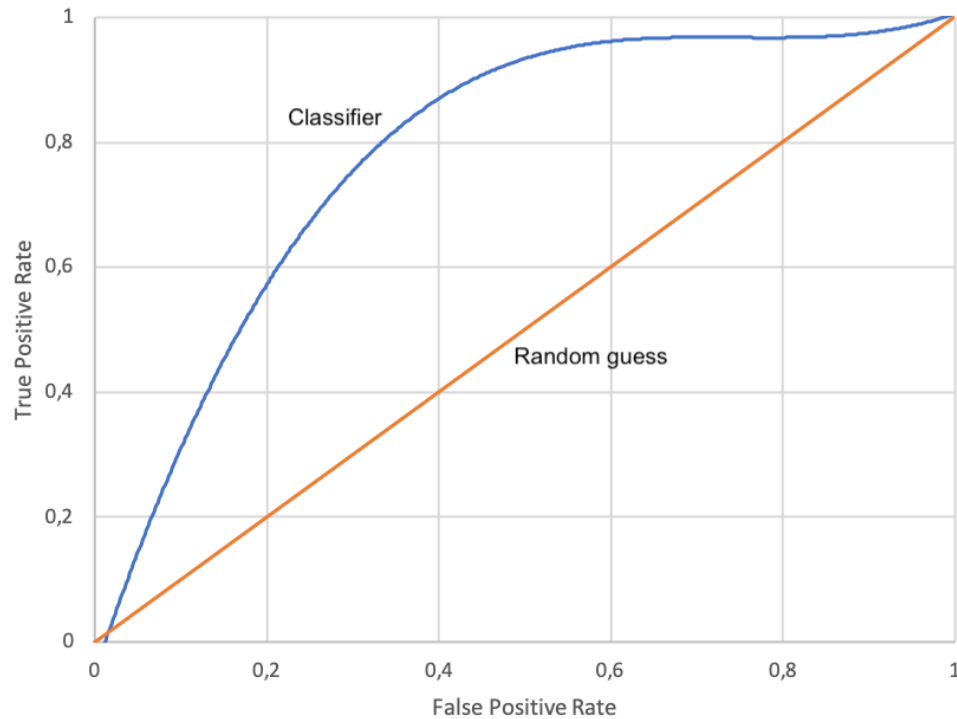
3.2 Data processing and algorithms

The data processing was conducted with R programming language (R Foundation 2019) and the “ChannelAttribution” R package developed by Davide Altomare and David Loris (Altomare & Loris 2019) for solving multi-channel attribution problem with Markov’s model. The package offers several functions for fitting a k-order Markovian model to marketing channel data and finding the correct order for the Markov’s model. Building a custom machine learning algorithm for fitting a Markov’s model to a dataset is a complex task and not in the scope of this thesis and therefore using a ready-made package is justified.

The package also offers tools for analyzing the model’s performance with Receiver Operating Characteristics (ROC) and Area Under the Curve (AUC). ROC and AUC are widely accepted tools for evaluating machine learning algorithm’s ability to classify the

data in to correct classes (Fawcett 2006). Machine learning algorithms such as Markov's model are typically classifiers which try to classify the data to 2 or more classes. In this case the algorithm tries to model whether the user will convert or not convert after the user has browsed through certain channel path. The conversion is perceived as the positive class and the non-conversion as the negative class. The user's path is given to the algorithm as the input. The algorithm can either classify the user as converted (positive class) or non-converted (negative class). By comparing the algorithms guess to the correct class the observation can be labeled as true positive, true negative, false positive or false negative. This four-field matrix is called the confusion matrix. By simulating a large amount of observations, we can calculate a True Positive Rate (TPR) and False Positive Rate (FPR). True positive rate describes how well the algorithm classifies the positive classes as positive and false positive rate describes the rate of observations labeled as positive even though they are really negative. These values are used to obtain a ROC curve. An example of a ROC curve is presented in the picture 6.

ROC analysis can be used to compare different classifiers' (algorithm classifying samples to different classes) performance in a classification problem. The TPR and FPR are plotted with various threshold levels and the ROC curves are compared to each other to find the most optimal algorithm settings. The upper left corner of the ROC curve (TPR = 1 and FPR = 0) represents the best possible classifier which classifies all the observations correctly. Accordingly, the right bottom corner (TPR = 0 and FPR = 1) represents the worst possible classifiers which classifies every observation incorrectly. However, this kind of classifier could be turned in to a good one by always taking a complement of the result. A linear line from (0,0) to (1,1) is equivalent to a random guess. As the classification results become better when the curve is closer to the left top corner, the comparison of the ROC curves is done by calculating the area under the curve (AUC). The AUC tells the probability of a positive observation to be classified as positive by the classifier. By comparing AUC numbers of different classifiers, we can tell which classifier works best for the classification problem in hand.



Picture 6. Example of ROC curves

The data processing begins by importing all the channel path data described above and converting all the values in right format for the algorithm. After this the minimum order for the Markov's model is approximated by maximizing the AUC value of the model. The ROC curve and AUC were estimated with minimum 100 points. The convergence parameter used was 0,05. This means the estimation process stops after the percentage of variation between the results of different simulations is less than 5%. The 6th order of the Markov's model had the best AUC value, so it was selected for the final attribution model. The AUC values of each Markov's model order are presented in the table 6.

Table 6 Markov's model's order AUC values

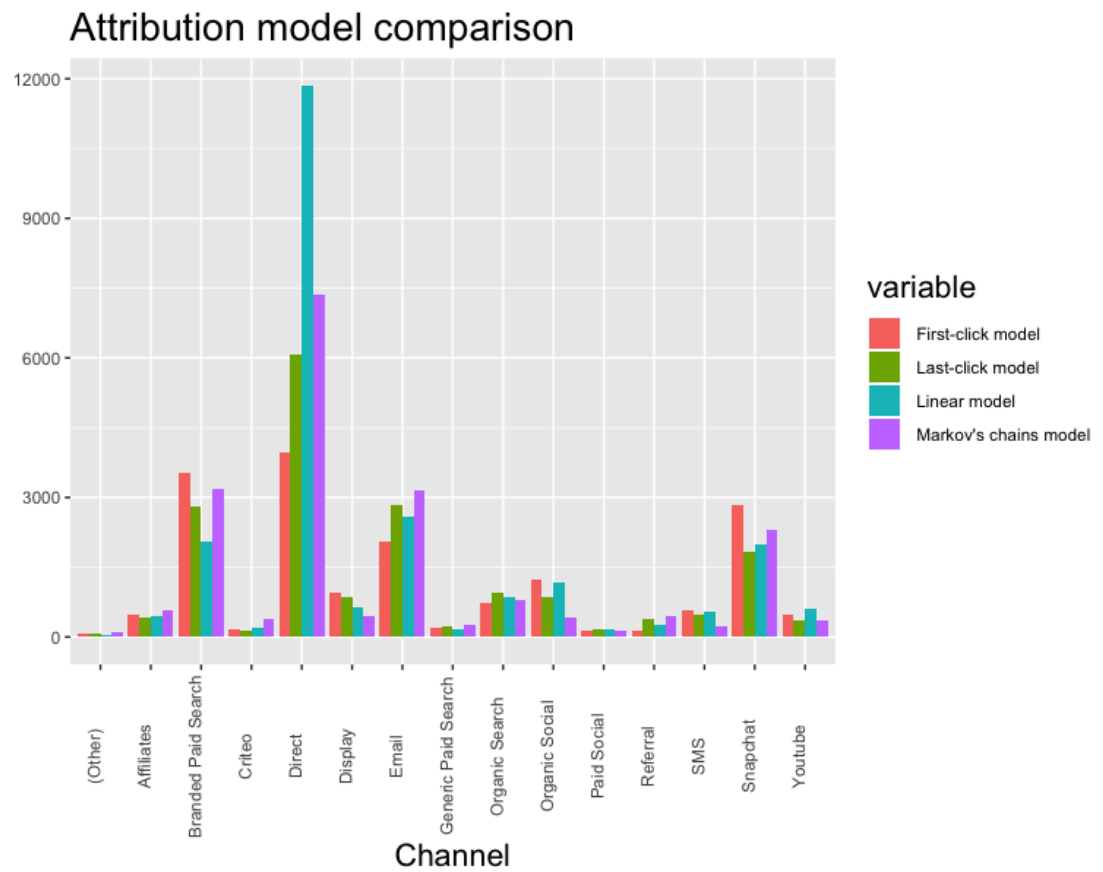
Model's order	AUC
1	0,615
2	0,635
3	0,646
4	0,665
5	0,685
6	0,705

After this the sixth order Markov's model was created and the attribution calculations were executed 10 times. The final results were calculated as the average values of the

10 calculations. The standard deviation between the results of the 10 calculations was also calculated and saved for further analysis.

The final results were collected in to a table that includes all the marketing channels and the amount of conversions and conversion value allocated to the channels by the Markov's attribution model. After this the same dataset is used to conduct a last-click attribution model analysis to see how a last-click attribution model would attribute the conversions and the value to the channels. These results are added to the results table, which now includes the amount of conversions and value attributed to the channels by the two attribution models. The table is then written as a csv file and moved to Microsoft Excel for further analysis and visualization.

The last-click attribution model was chosen as the main comparison model for the Markov's attribution model because the case company had previously used this model to allocate their marketing budgets to the channels. However, the analysis was also conducted using other heuristic models to see what kind of results they give and how much differences are there between the different channels. The same dataset was used to conduct a heuristic attribution model analysis using first-click, last-click linear, and time-decay models. The results were plotted in to a graph using the ggplot function in R studio. An example graph of the heuristic attribution model comparison is presented in the picture 7. The numbers in the graph are changed due confidentially. The results were also written as a csv file for further analysis in Microsoft excel. In Excel all the heuristic models and the Markov's attribution model were compared to the average results of all the models to see how much they differ from the other attribution models in absolute and percentual values.



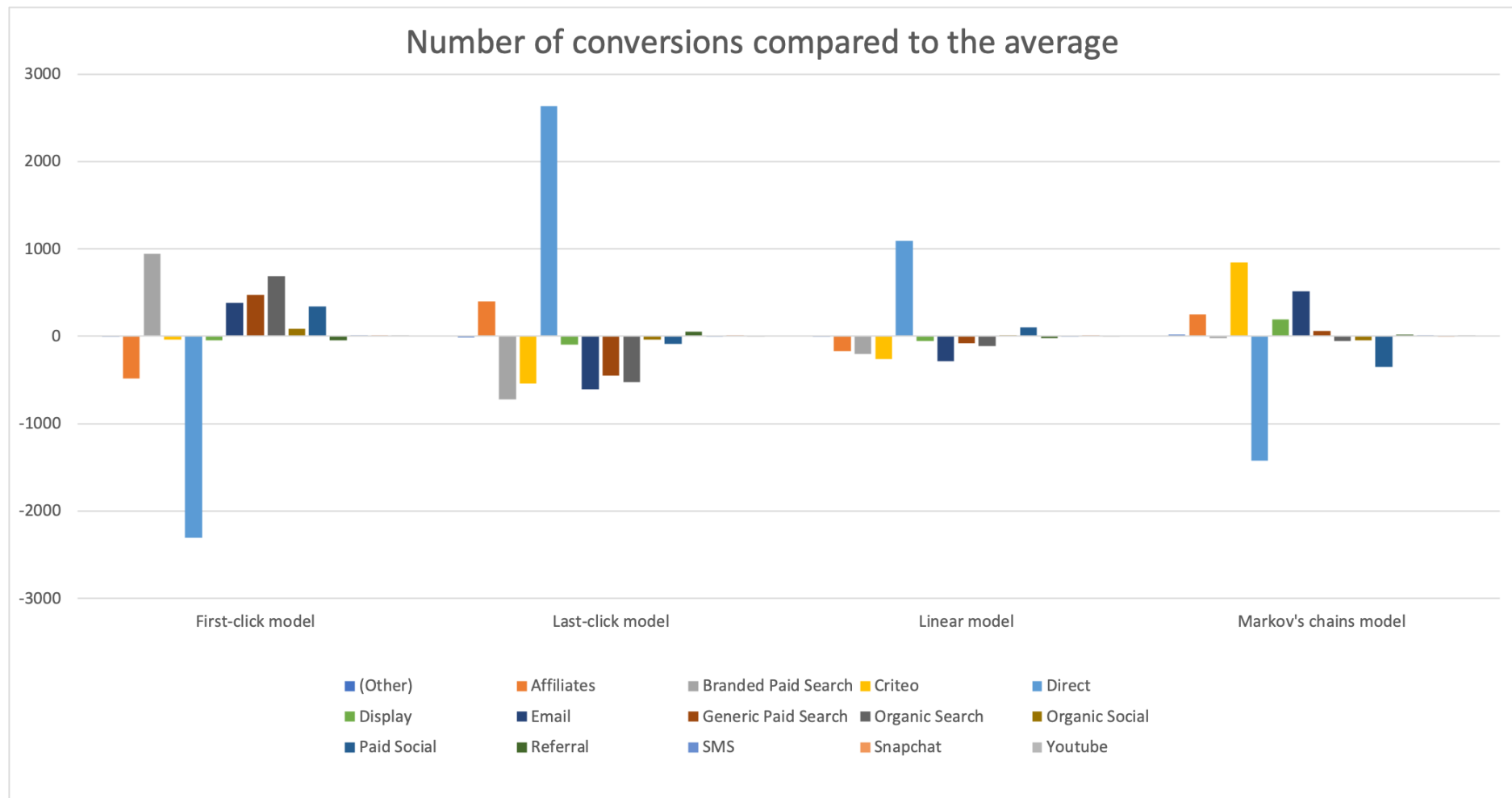
Picture 7. As example of a heuristic attribution model comparison graph drawn with R studio

The Markov's model was also used to build a Markov's graph that illustrates the transitioning probabilities between the different channels. The graph was made using the transitioning matrix and the igraph package in the R coding language.

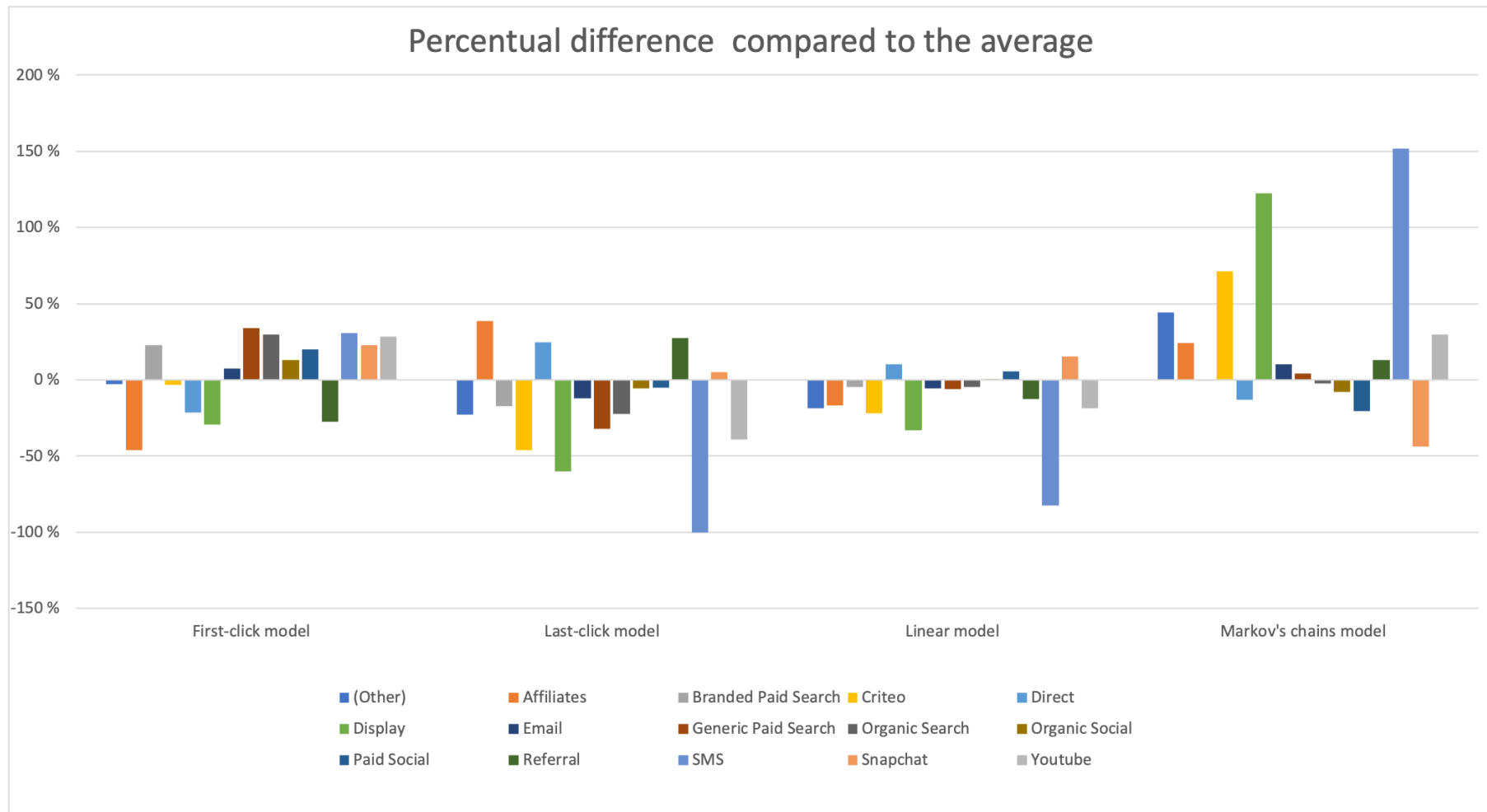
4. RESULTS

The amount of conversion attributed to each of the channels by different attribution models are presented in the picture 13 in the appendix A. Markov's chain attribution model was compared to 3 most traditional heuristic attribution models that are most used in digital marketing industry. The differences in the conversions attributed to each of the channels vary highly between the different attribution models depending on their characteristics and how they attribute the value to the channels. The differences are best illustrated by comparing the different attribution models to the average number of conversions given by all the models. This comparison is presented in the picture 8.

As explained in the chapter 2.2.1, the first click model favors the first channels in the marketing funnel because it gives the whole credit to the first touch point in the funnel. The last-click model does the opposite by giving the full credit to the last touch point in the funnel. The linear model favors all the channels in the funnel with equal priority. The Markov's model is not based on any rule based on the channels position in the funnel but instead it models the funnel based on probabilities and favors the channels that increase the probability to convert the most. Based on these assumptions we can conclude that Direct channel seems to usually be part of the last steps in the funnel as the last-click model favors it much more compared to the average and the first-click model less than the average. Branded paid search, generic paid search, email and organic search have opposite results which means they are more often located in the first steps of the funnel. The linear model gives quite average results from all the attribution models. The Markov's attribution model gives very different results than any of the heuristic models because it favors some channels like the Criteo and Email much more than any of the other attribution models. The Direct channel is favored less than the attribution models in average.



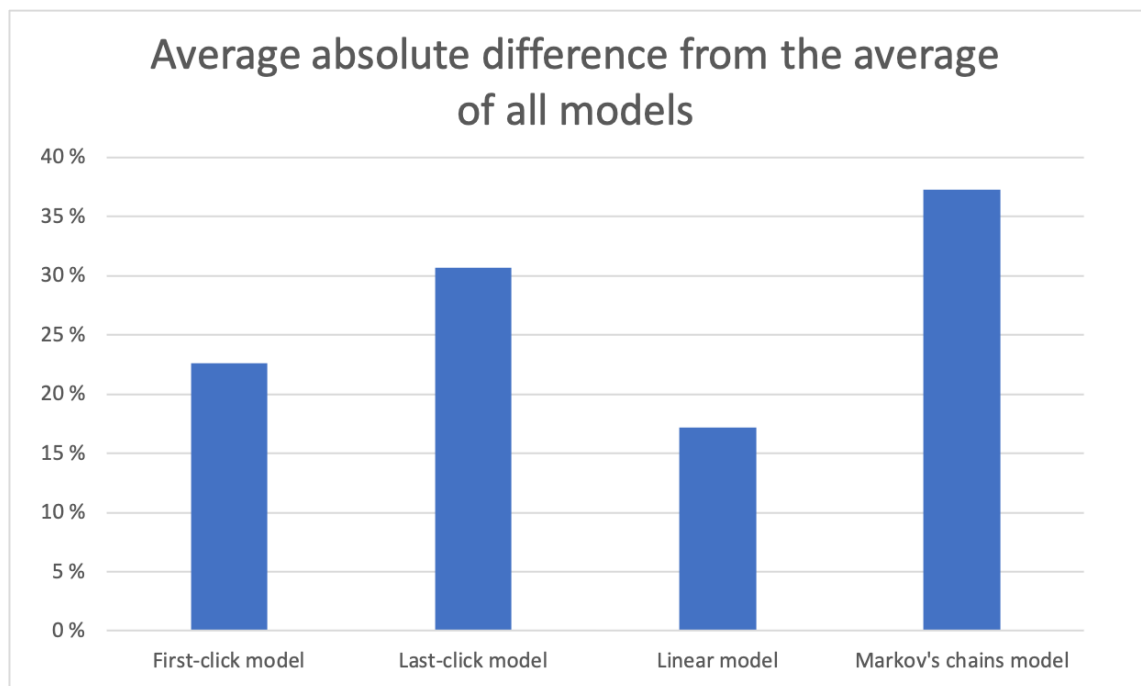
Picture 8. Number of conversions compared to the average of the attribution models



Picture 9. The percentual difference of the value attributed to the channels by each of the channels compared to the average of all attribution models

The same kind of conclusion can be drawn from the picture 9, which presents the percentual difference of the attribution models compared to the average of all attribution models. From this chart it is easy to notice that the Markov's attribution model gives very different values for some of the channels than the other attribution models. Especially the SMS and Display channels are valued much more than the average of the models values them. The amount of data for the channel SMS is very small and even small differences can cause large percentual difference, but the Display channel has much more data and the large percentual difference means that the Markov's model attributes a lot more value to it than the other channels.

The Markov's model's difference compared to other attribution models can also be confirmed by looking at the average absolute difference of the attribution models from the average of all models. These values are presented in the picture 10. From this comparison we can see that on average the Markov's attribution model differs 37% from the average of the all models and the difference is larger than for any other attribution model.



Picture 10. The Average absolute difference from the average of all models

Comparing the results of the new attribution model gives a good idea on how the attribution model favors different channels compared to other models. However, the optimal way of attributing the conversions to the channels can't be proven by comparing different models against each other. Instead, the model's accuracy must be evaluated with the model's ability to predict the conversions. The model's accuracy will be discussed in more detail in the chapter 4.2.

In this thesis we focus on comparing the results of the Markov's attribution model to the last-click model, because it is the most used attribution model in the industry and also the case company has been using it before utilizing the Markov's attribution model. The performance of the marketing channels can be either evaluated based on the amount of conversions or the amount of conversion value (revenue) each of the channels bring to the business. In this thesis we focus on the amount of conversion value, because the amount of conversions itself does not bring as much value to the business as the actual revenue that comes from the channel. Some channels can drive more valuable customers that purchase more at once and therefore the conversion amount is more valuable metric in this case. The table 7 presents the differences in the value of the conversions attributed to each of the marketing channels by the Markov's attribution model and the last-click model. A positive value difference means, that the Markov's attribution model attributed more conversion value to the channels than the last-click model. A negative value means that the Markov's attribution model attributed less conversion value to the channel than the last-click model.

Table 7 Difference in the conversion value between Markov's attribution model and last-click attribution model

Channel	Value difference %
Criteo	164 %
Organic Search	47 %
Email	21 %
Generic Paid Search	66 %
Branded Paid Search	14 %
Display	457 %
Other	183 %
Referral	22 %
Youtube	211 %
SMS	0 %
Snapchat	-10 %
Organic Social	-12 %
Affiliates	-7 %
Paid Social	-7 %
Direct	-29 %

From the table 7 we can see, that the Markov's attribution model attributed much more conversion value to the following channels: Criteo, Organic Search, Email, Generic Paid Search, Display, Other, Referral and Youtube. This means that the Markov's attribution model suggests the company should invest more budget to these channels than what last-click model suggests. As the former budget allocation decisions were made based on the last-click attribution model, the company should re-evaluate the budget allocation with the new knowledge provided by the Markov's attribution model. On the contrary, the

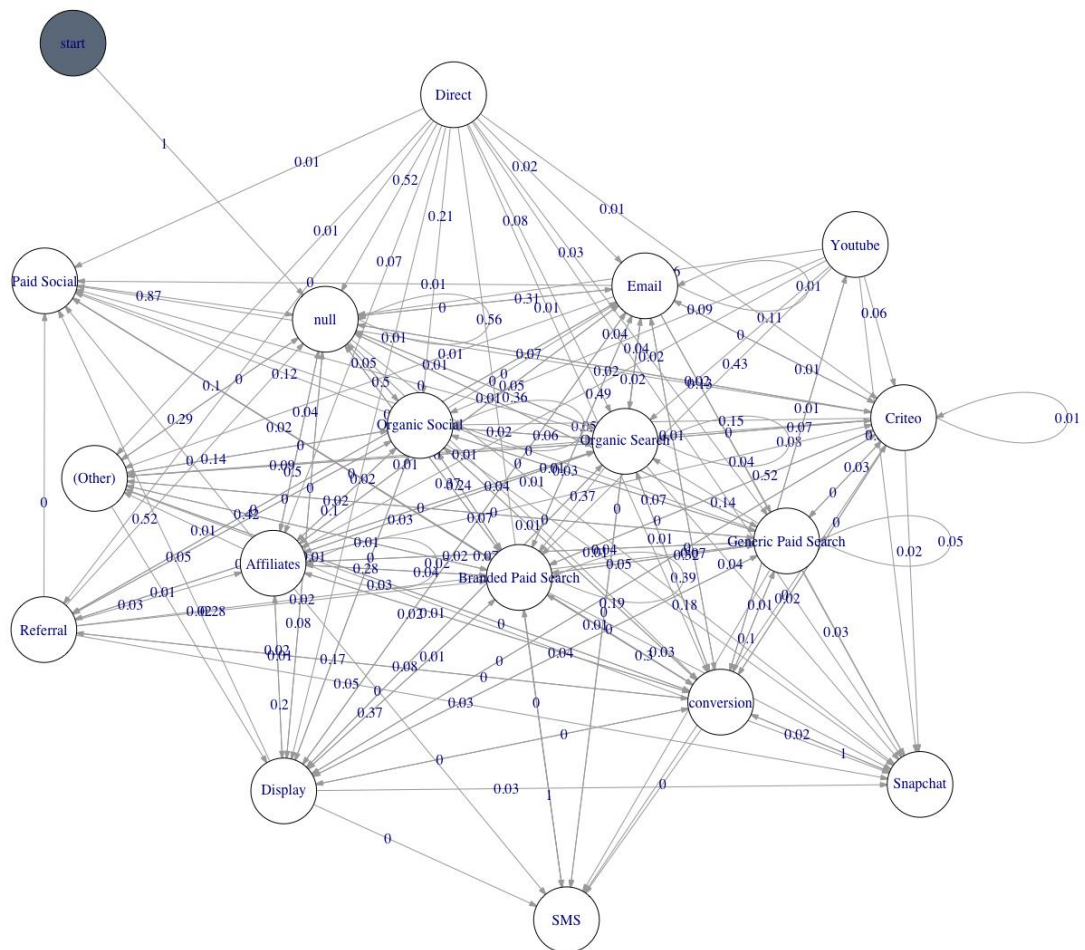
Markov's attribution model allocates less conversion value to the following channels: Snapchat, Organic Social, Affiliates, Paid Social and Direct. This suggest that the company should invest less budget to the channels than they did before using the last-click attribution model.

Although the percentual conversion value changed quite drastically for some channels, the channels need to be also analyzed in terms of absolute value they bring to the company and the marketing channels ability to scale with larger budgets. Even if the value of the channel changes lot percentually, it may still be a very small channel in absolute numbers and therefore the channel is not that important to the company. The channel can also have already very high budgets compared to the overall size or audience in the channel and therefore it may not be possible to scale the budget in though the attribution model would suggest so. The table 10 in the appendix A presents the percentage of each channel's conversion value compared to the total conversion value of all channels according to the Markov's attribution model. From this table we can see that the Direct, Branded Paid Search and Email are driving majority of the total conversion value. This suggests, that these channels have most potential in terms of conversion value and even a small percentual change in the allocation of the conversion value can have a large impact the business if the budget is allocated in more optimal way with the new attribution model. On the other hand, the channels Youtube, SMS, Snapchat and Other are all driving less than 1% off the total conversion value. Therefore, the changes in amount of conversion value attributed to these channels does not have that much impact to the budget allocation even though the change would be large percentually. Each of the channels and their budget allocation are further discussed in the chapter 5.

The detailed results of the attribution modeling analysis are presented in the table 9 in the appendix A. The table presents the amount of conversions and the conversion value for the Markov's attribution model and the last-click attribution model. The differences between the amount of conversions and the conversion value are also presented in the table. The table is sorted in a descending order by the conversion difference. Higher value in the conversion difference means that the Markov's model attributes more conversions to the channel than the traditional last-click model. The conversion difference and the conversion value difference columns are color scaled to improve the readability of the table. The results and their implications to the budget allocation are further analyzed in the chapter 5.

4.1 Transition probabilities

The Markov's model also allows us to see the calculated transitioning probabilities between the channels. Transitioning probability means the probability of a user moving from one channel to another. These probabilities are collected into a transitioning matrix. As the study contains large number of different marketing channels the transitioning matrix is a 18x18 matrix with all the marketing channels and the start, null and conversion states to illustrate the start of the funnel, the customer journey that does not end in conversion and the customer journey that ends in conversion.



Picture 11. Markov's chain's transition probabilities graph (values under 0,01 removed for clarity)

The transitioning matrix can be plotted as a graphical illustration of the transition probabilities called a markov's graph using the 'markovchain' R coding language package (Spedicato et al. 2016). The Markov's graph of the dataset used in this study is presented in the picture 11. The transitioning probabilities under 0,01 are removed from the graph for clarity.

The transitioning matrix tells us for example, that a user coming from the Criteo channel has a 12% probability to come back to the site from Facebook. This is a very high transitioning probability and most of the values are much lower than that. The marketing team can utilize these values for campaign planning to identify channels that usually work together to drive conversions and try to unify the customer journey and make the brand experience as unified as possible especially in these channels.

4.2 Removal effects

Markov's model also allows us to see how much removing a single marketing channel would affect on the total amount of conversions and conversion value. The Removal effects of 6th order Markov's model are presented in appendix A, picture 14. The more detailed results are also presented in the appendix A, table 11. The removal effects for higher order Markov's models are calculated as the mean of all the removal effects of the sequences that include the marketing channel. A higher removal effect means that the channel was part of many sequences that led to conversion and therefore the channel is important for the business. For example, the Direct channel has a removal effect of 60,7% to the number of conversions. This means that the Direct channel was part of 60,7% of all the customer journeys that ended up in conversion. The number of conversions attributed to the Direct channel is calculated by dividing the Direct channel's removal effect with the sum of all removal effects.

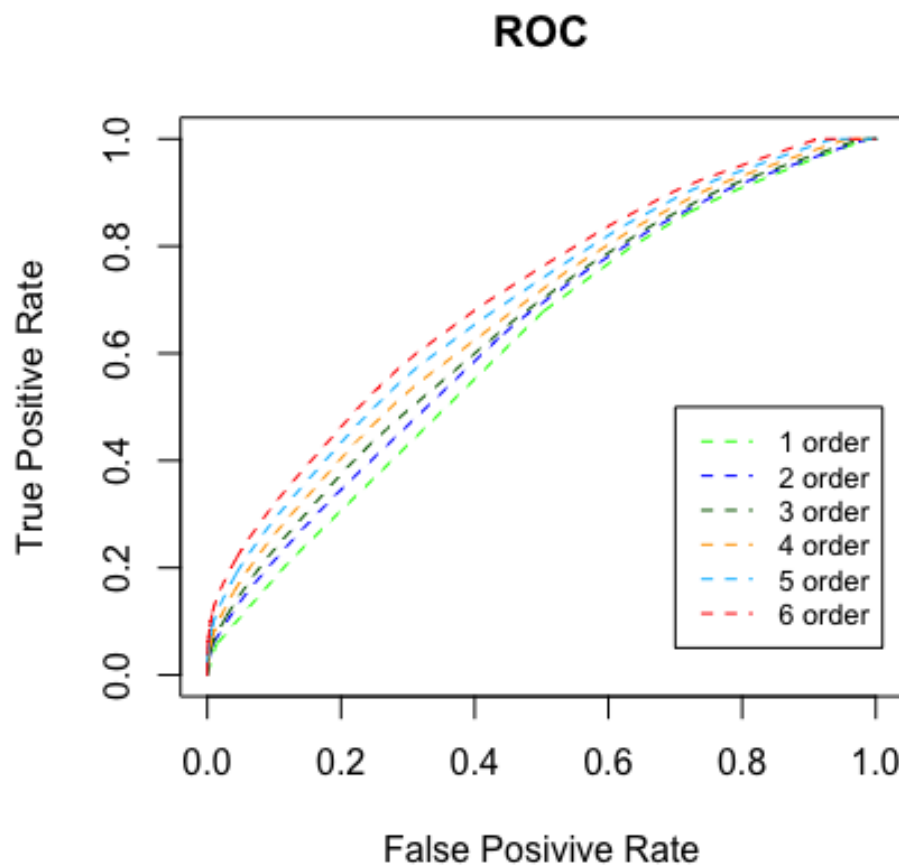
From the picture 14 we can see, that also Email, Branded Paid Search and Organic Search have very high removal effect. All of these channels are used by existing customers to enter the site either by using the website address field, search engine results or by clicking a newsletter that they have subscribed to. Therefore, it is not a surprise that these channels are part of many customer journeys and therefore very important for the business.

4.3 Prediction accuracy and model reliability

As argued by Anderl et al. (Anderl et al. 2014) and Shao & Li (Shao & Li, 2011), the prediction accuracy of the attribution model is important measure that indicates the model's ability to predict where conversions are coming from in order to be reliable when planning marketing budgets. Another important factor is the model's ability to give robust e.g. stable and reproducible results so that the results do not have too much variance. Both of these factors are important in terms of research but also if the attribution model is used in managerial decision making.

As described in the chapter 3.2, the prediction accuracy of the model can be measured with ROC curve and the AUC value. The ROC curves of 1-6 order Markov's models are presented in the picture 12. The ROC for the 6 order Markov model is marked with red dashed line. As we can see from the picture and the table 6 in the chapter 3.2, the area under the ROC curve (AUC) is highest for the 6th order Markov's model, which means the prediction accuracy of the model is the highest. The AUC value of 0,705 is not usually considered very good as it means the model can predict 70,5% of the samples correctly. However, the accuracy and its requirements vary based on the application and the data used and for an attribution model we can accept far worse AUC values than for example is the application is in medicine research.

For example, Alblas (Alblas, 2018) was able to get as high as 0,9043 AUC value in his study by using 5th order Markov's attribution model. The number of marketing channels in the study was smaller which could explain the much better results of the study.



Picture 12. The ROC curves of 1-6 order Markov's models

The last-click attribution model that is currently used by the company and also has wide acceptance in the industry is not officially a model that is used to predict any conversions. However, to compare the results of the attribution model chosen for this thesis, we can

estimate the prediction accuracy of the last-click model by assuming that it would behave in a same manner as the first order Markov model. The first order Markov model and the last-click model both only look at the previous step before the conversion. Therefore, both of the models only depend on the conversion rate of the channel to make the decision on whether the user will convert or not. The models don't care which channels the user has used before arriving to the last state before the conversion. As we can see from the table 6 the AUC of the first order Markov's model is 0,615 and we can use the same AUC value for the last-click model. This indicates that the Markov's 6th order attribution model is better at predicting user behavior than the last-click model and therefore this model is better than the current attribution model used by the case company.

The robustness of the model was evaluated by calculating the average standard deviation of the number of conversion and conversion value in each of the attribution calculations. The average standard deviations are presented in the table 8. We can see from the table that the standard deviation is very high for Snapchat and SMS channels. This is caused by the small amount of data for those channels. For other channels the standard deviation is rather low and the total standard deviations for the model are 2,93% for the conversions and 3,90% for the conversion value. This proves the model's robustness and that it can be used to make managerial decisions about budget allocation.

Table 8 Standard deviations of the attribution model results

	Average standard deviation	
	Conversions	Conversion value
Paid Social	0.77%	0.73%
Direct	0.11%	0.19%
Branded Paid Search	0.30%	0.31%
Organic Search	0.44%	0.54%
Generic Paid Search	0.72%	0.68%
Email	0.19%	0.40%
Organic Social	1.11%	1.51%
Criteo	0.40%	0.66%
Affiliates	0.62%	1.05%
Snapchat	11.53%	11.49%
Youtube	7.26%	10.16%
Referral	1.05%	2.33%
Display	1.77%	2.13%
(Other)	2.45%	5.13%
SMS	15.25%	21.18%
Total	2.93%	3.90%

The prediction accuracy and robustness of the model could be further improved by decreasing the number of channels and removing the channels that belong to only small

amount of user paths. Also increasing the data collection timeframe could improve the results as there would be more data for the model to learn from.

5. DISCUSSION

The results of the attribution model analysis are not as easy to interpret and put in practice as one could expect from the initial results that show very clearly which channel is better than the other one. The reason for this is that the company may not be able to effect on all of the marketing channels and their growth by allocating more time or budget to the channel. The attribution model also doesn't consider the current cost of the advertisement in the channels and therefore the results must be further analyzed by taking in to account the budgets, time and scaling possibility of the channels. In this chapter the results of the attribution model analysis and their implication to the company and marketing budget allocation are discussed in more detail

The results are also compared to a similar study conducted by Alblas (2018) in the chapter 5.2 where the changes in the value of conversions are discussed. Although the attribution model used in the study (Markov's attribution model) was same as in this thesis, the results and the prediction accuracy of the attribution model can vary due the differences in the datasets used in the analysis. As mentioned in the chapter 4.2, the AUC value in the Alblas' study was significantly higher which indicates a better prediction accuracy and therefore a better attribution model. However, it is encouraging that the results in both studies support each other and the changes in the conversion value are aligned for most of the channels included in both of the studies.

5.1 The value of marketing channels according to the Markov's attribution model

According to the Markov's attribution model the Direct channel is driving a substantial amount of the total revenue. The result is easy to understand as many people coming directly to the website are already looking for something specific to buy from that website and therefore, they convert very well compared to some traffic from ads which tend to drive more people that are just checking out the offer and don't yet know the business that well.

Another source of well converting traffic that contains many existing customers is the Email channel. The Email is the second valuable channel according to the Markov's attribution model. This channel is one of the main advertisement channels for the company and therefore the result is not surprising. The Email channels cannot be scaled just by allocating more marketing budget to the channel, but the company can, for example,

invest more time to make the email marketing even better with better audience segmentation and more personalized content.

In the third place is Branded Paid Search which is also very similar channel than the Direct channel. In the Direct channel the customer uses the browser's address bar or a bookmark to navigate to the page. In the Branded Paid Search, they use the search engines search field to search for the business' name and click the paid advertisement to get to the page. In both of the cases the customer already knows the name of the business and probably their offering too. It is not surprising that these channels perform very well according to any attribution model. The problem with these channels is, that these are not channels that the company can scale by adding more budgets or resources to these channels. Therefore, we focus more on the other channels.

The fourth most important channel is Organic Search which contains both brand keyword searches and other keyword searches such as the customer searching for a specific product. The amount of brand keyword searches is usually driven by the brand awareness and other advertisement which cannot be scaled using resources to the Organic Search channel. However, the amount of traffic from other keyword searches can be increased by using resources to improve the search engine optimization of the page. This can mean, for example, that the company adds more product information to the page or increases the site loading speed to make the search engines rank the pages higher in the searches. However, as most of the traffic in Organic Search channel comes from the brand keyword, scaling channel significantly is probably not very feasible.

The other channels in the analysis were all attributed less than 6% of the total value. The value was concentrated on the first four channels, which is quite surprising to see as all of these channels are quite hard to scale as they are not depended on the marketing budget but rather the overall brand awareness and number of existing customers. The traffic coming from these channels is also very depended on the paid advertisement channels which have built the awareness in the first place and also act as a reminder for the people to come and visit the site. The next chapter will concentrate more on these paid channels as the budgets are currently divided based on the last-click attribution model and we are interested to find out which channels should be allocated more marketing budget.

5.2 Comparing results of Markov's attribution model to the last-click model

Comparing the results between the two attribution models is done by comparing the amount of conversion value the models attribute to each channel. The results of the

attribution analysis are presented in the appendix A in the table 9. The difference in value is calculated to the second last column of the table and color coded to improve readability. The results are analyzed from the bottom to the top to see which channels lost most value and which channels gained the most value.

From the results we can see that the largest absolute change in the conversion value is in the Direct channel. According to the Markov's attribution model the value of the channel is 29% smaller than last-click model indicates. Percentually this change is smaller than for many other channels, but as we saw in the previous chapter, the value of Direct channel is major part of the total value and therefore there is much absolute value to attribute. As the amount of value for Direct channel decreases, there's now more value to divide to the other channels. The study of Alblas (2018) indicated a similar decrease in the conversion value for the Direct channel.

We can also see from the results, that there are also few other channels which value decreased according to the Markov's attribution model. Snapchat, Organic Social, Affiliates and Paid Social have all lost some of the value when comparing the results of the two attribution models. However, the changes are only 7-12% and therefore it won't have much effect on the overall allocation of the budgets.

The SMS channel has 0 value in both of the models. There hasn't been any major traffic from the channel in the analysis time period and therefore there's also no last-click conversions or conversion value. However, the channel has been part of couple of customer journeys that have led to conversions and therefore Markov's attribution model attributes a very small value of conversions and conversion value to this channel in some of the simulation runs. The average value between the ten simulation runs is 0, but as we can see from the chapter 4.2, the variation of the results is not 0.

The conversion value of YouTube, Referral and Other channels is quite small in absolute value and therefore the absolute changes in the value are also quite small even though percentually the value of the channels has increased quite much. Therefore, we can conclude that these channels are not that important for the company in terms of driving conversions and conversion value.

The Markov's attribution model gives Display channel 457% more value than the last-click model. Although the absolute value of the channel has been quite small in terms of last-click model, the value now has become more significant. The change in the conversion value of this channel was percentually largest of all channels.

The value of all the search engine related channels increased significantly in the comparison. The value of the Branded paid search increased 14%, the Generic paid search

66% and the Organic search 47%. Especially the change in the value of the Generic paid search is very interesting, because the increase is so large, and the channel offers the company a possibility to scale the budgets. Also, the change in the value of the Organic search can be taken into account in search engine optimization resourcing. The study of Alblas (2018) indicated similar increase in the conversion value for the generic paid search. However, in Alblas' study the value of conversions in the branded paid search decreased which is contrary to the results in this thesis. The difference can be caused by differences in the datasets and the customer behavior of the companies analyzed.

The Email channel has the largest absolute value after the Direct channel according to the both attribution models. The Markov's attribution model gives Email 21% more value than the last-click model. This result is also in line with case company's expectations, because the Email channel has always seemed to give more actual lift to sales than the last-click attribution model metrics have shown. The study of Alblas (2018) indicates similar increase for the conversion value of the Email channel.

The largest increase in absolute value goes for the Criteo channel. The conversion value has increased 164% and the absolute value of the channel is now fifth largest of all channels. The change in the value of this channel is very considerable and should be taken in to account in the marketing budget allocation.

5.3 Recommendations of budget reallocation

As described in the chapter 5.2, the importance of Direct, Paid social, Affiliate and Organic social channels has decreased according to the Markov's attribution model. The Direct and Organic social are non-paid channels and therefore we can't decrease the budget in them. However, we can move some of the budget in the Paid social and Affiliate marketing to channels which importance grew according to the attribution model analysis.

The marketing channels that gained most value according to the Markov's attribution model were Criteo, Organic Search, Email and Generic paid search. As mentioned before, the organic search and Email are not scalable with budgets. However, Criteo and Generic paid search have marketing budgets and the attribution modeling analysis indicates that the company should allocate more marketing budget to these channels than they did before to gain better results. Especially the value of the Criteo channel has grown very significantly and it should be emphasized in terms of budgeted and time spend optimizing and making content to the channel.

Even though the absolute value change of Display marketing channels wasn't as large as for some of the channels, the percentual change was very significant. The total value of the channel is quite moderate compared to the other channels, but the Markov's model indicates that there could be a lot more potential in the channel as the last-click attribution model has indicated in the past. Therefore, the company should invest some resources to further test the Display channel and see if the channel can be scaled with good results.

5.4 Roadmap for utilizing the attribution model in the case company

Online store management is a fast-paced business where traffic and marketing channels are tracked every day and real time data is very important because it allows the company to react fast when customer behavior changes for some reason. The purpose of this thesis was to compare the marketing channels and their performance in a long time period, and it gives a good indication on how the channels have been performing within longer period of time. However, sometimes the performance of the channels can change quite fast and the company should pursue for faster analysis cycle. Especially when the amount of data continues to rise, it is possible to use the Markov's attribution model as a daily or weekly decision-making tool.

After the case company has implemented the recommended budget allocations, they should monitor the results carefully and continue attribution analysis as a weekly process. The results between the channels continue to change all the time depending on the creatives, the price of advertisement in each channel and changing customer behavior. Therefore, the attribution analysis is not an analysis that can be implemented just one time but rather a continuing process of how the digital advertisement results are tracked. Now that the company has done the analysis once it is easy to repeat the analysis using the same code base and analysis methods. The results should be always used to conduct new budgeting allocation decisions and the process should aim to keep the budget allocation as optimal as possible all the time.

It would be even possible to automate the whole attribution modeling process quite easily and integrate the Markov's attribution model to the real time dashboards the case company is using to keep track on business metrics and online marketing. Markov's attribution model is not a computation intensive task and therefore the model could be recalculated once a day using for example the last 30 days of data. After that each channel could be given a multiplier that is used during the day to allocate the traffic and the conversions to the channels to see which channels are driving the performance during the

day. With shorter data collection time the attribution model can react faster to the changing business environment and customer behavior, but there's a risk that the amount of data is not sufficient, and the variance of the results grows too much. However, it is easy to test the sufficient amount of data using the methods in this thesis to test how many days are needed to collect enough data for robust analysis with small variance and high prediction accuracy. As the amount of data continues to grow when the company scales, the data collection period can be shortened, and the model can use the latest data to give marketing budget allocation decisions.

5.5 Limitations of the study

As mentioned in the chapter 3, the dataset used in this research includes only online advertisement and clickstream data. Therefore, any offline marketing will not be included in the analysis. In this case the company hasn't done any significant offline advertisement during the dataset collection, which is why the offline advertisement does not affect the results. However, if the same analysis would be conducted to a different company and dataset, the offline channels and their impact on the results should be notified. The offline marketing isn't as easy to measure as online channels and it would probably improve the results of some channels more than others, and therefore a reliable analysis would be harder to conduct for a company that uses both online and offline marketing.

The dataset used in this research only contains clickstream data and doesn't take in to account the impressions of the advertisement that didn't lead to click and a website visit. Collecting impression data on all of the marketing channels used in this research is not feasible due limitations in some of the advertisement channels and the resources of the company.

The attribution analysis also doesn't take into account the current budgets of the marketing channels and the ability to scale them. The results indicate the relative importance and value of each channel, but it doesn't tell which channel's budget can actually be increased. This decision will be left for the managers to make. Some channels may be able to give more results when the budgets are scaled, but in some channels the performance may already be achieved and increasing the budget just increases the spend but does not give more conversions or value. The recommendations of budget reallocation must therefore be tested in action and after the test the attribution analysis can be conducted again to see did the results improve or not. By iterating many tests and analysis the company can achieve more optimal budget allocation between the marketing channels.

5.6 Topics for future research

The future research in this subject should focus to improve the reliability of the analysis and to improve the understanding of the impact of budget reallocation. The attribution analysis conducted in this research doesn't take in to account the budgets of the channels and the budgets relative size compared to the marketing channel's size and reach. Therefore the attribution analysis can indicate that the company should invest more money to a channel even though the channel is already fully exploited and the budget change would not reach any new customers and it would only increase the frequency of the ads seen by the users to a level that is not optimal. It is common in digital marketing that the price for impression rises when the company increases their budget because the company also has to bid higher to win more ad auctions to show the ad to more people. At first the platform's optimization algorithm can choose the cheapest impressions but as the budget gets larger it has to move to more expensive impressions and therefore the results get worse. For this reason, understanding the actual effect the budget change can have in each of the channels' performance would improve the practical usability of the analysis. This is especially important in smaller markets such as Finland where the reach of the marketing channels is quite limited due small population and it is quite easy to reach too high frequencies even when targeting everyone in the country. For larger markets this is not as important as it is usually easier to scale the marketing spend in one channel longer as there are so many people in the market. Therefore, the price of the advertisement does not rise as fast when scaling budgets and also the frequency for one user does not rise as easily, which makes scaling budgets easier. There are some tools available for predicting the rise of the costs when scaling budgets. For example, Facebook and Google both offer tools called reach planner which can be used to model the prices when the reach of the campaign gets higher with larger budgets. However, including these metrics to the attribution analysis is much more complex task and requires further research.

As mentioned in the previous chapter, this research is only using clickstream data. The reliability of the attribution analysis could be improved by taking into account also the ad impressions and therefore the future research should investigate how much the results of the analysis change if the dataset would also contain impression data for individual ads. In this case the company would know which ads every customer has seen before entering the site through any channel. For example, the customer might have seen an ad in social media and then entered the site using a search engine's paid ad. In this case the clickstream data gives wrong kind of indication on which channel should get the credit

for the conversion and the impression data could help to solve that issue. The data collection can be conducted with impression tags which is basically just an URL that the browser calls when the ad is shown to the user in the ad platform. The URL contains information such as the campaign and creative name so the impression tagging system knows which kind of ad was shown to the user. The impression tag also places a cookie to the user's browser so next time the user visits the website the impression tagging system knows which ads were shown to this specific user and this can be used to analyze the effectivity of the advertisement. Most of the modern marketing channels support impression tagging already, but some channels don't yet support these impression tags. However, these channels could be excluded from the analysis. The ad impressions could be counted as touchpoints in the Markov's attribution analysis so the same methodology could be used in the analysis.

The current attribution model analysis doesn't consider the individual ad or the campaign that the customer saw and clicked. The results of advertisement may vary drastically between different campaigns, ads, landing pages and products advertised and therefore reliability of the analysis could be improved by taking more data into account. However, if the analysis would be conducted with, for example, campaign data instead of marketing channels data, the requirements for the amount of data would grow drastically. This kind of analysis would not be possible for a small or even medium sized companies as the amount of data for each campaign would be too small and the variance of the results would grow very large. For larger companies running advertisement campaigns with hundreds of thousands the amount of data for individual campaign would be enough to make the attribution analysis in campaign level and find out which individual campaigns are more valuable compared to others. For example, when companies launch multiple new campaigns in many channels at the same time it is hard to tell which campaigns are working better than the others especially if the attribution analysis is made in channel level. The analysis in individual campaign level would improve the company's ability to make faster budget allocation decisions based on the current campaigns' performance.

Combining more data to the analysis could also improve the practical use of the results as the results and their true effect for the company would be easier to interpret. For example, instead of measuring the revenue of each channel, the conversions and their value could be analyzed in actual profit the conversion has generated for the company. Measuring the profit for individual purchase may not be feasible for all kinds of business' but at least most e-commerce companies can easily calculate the gross profit for each conversion in real time and use this data to determine the marketing channels that drive more actual profit. These channels could be emphasized in the marketing mix more than

the channels that drive low profit conversions. This would help the company to drive for better profitability.

6. CONCLUSION

The importance of online marketing has grown radically in the recent decades and the companies all over the world are investing larger portion of their advertisement budgets to online channels every year. The online marketing is known as easily measurable marketing which real affect and performance can be tracked in detail. Despite of common belief, the measurement and performance analysis of online marketing is far from simple as it is usually conducted in multiple channels which results depend on each other and the results of the performance analysis vary drastically depending on the attribution model used. This thesis aims to answer to the research question: “How to make optimal digital marketing channel budgeting allocations?”, by conducting a data-driven attribution model analysis to the case company’s dataset and comparing the results with the de-facto last-click attribution model’s results. The new attribution model is based on a k-order Markov’s chain.

The thesis begins with literature review to online marketing, measurement techniques and most used attribution modeling models in the industry. The Markov’s attribution model was chosen to the analysis because of its promising results in other research and the ease of implementation with the dataset available. The dataset used in the analysis contains 582 111 user paths collected during 7 months period from the case company’s website. The analysis was conducted using R programming language and open source ChannelAttribution package that includes tools for fitting a k-order Markovian model in to a dataset and analyzing the results and the model’s reliability. The performance of the attribution model was analyzed using a ROC curve to evaluate the prediction accuracy of the model.

The results of the research indicate the Markov’s model gives more reliable results on where to allocate the marketing budget than then last-click attribution model that is widely used in the industry. The results of the Markov’s model attribution were compared to the last-click model and the results indicated that the company should invest more budget on Criteo, Generic paid search and less budget on the Paid social and Affiliate marketing channels. Also, as there is no marketing budget in the Email and the Organic search marketing channels, the company should use more time to improve the results in these channels. The Display channel also showed promising results and the company should further investigate the potential of the channel by testing. The results were also compared to a similar study conducted by Alblas (2018) using same attribution modeling

technique. The results were mostly in line with this thesis which supports the reliability of the results.

Overall the objectives of this thesis were achieved, and this study provides a solid framework for marketing managers to analyze their marketing efforts and reallocate their marketing budgets in more optimal way. However, more research is needed to improve the prediction accuracy of the model and to improve the understanding of the effects of budget reallocation.

REFERENCES

- Abrams, Z. and Vee, E., 2007, December. Personalized ad delivery when ads fatigue: An approximation algorithm. In International Workshop on Web and Internet Economics (pp. 535-540). Springer, Berlin, Heidelberg.
- Alblas, R. (2018). Attribution modeling. Using conversion value as an alternative attribution measure to understand the customer journey online, Trauner,
- Altomare, D. & Loris, D. ChannelAttribution <https://cran.r-project.org/web/packages/ChannelAttribution/ChannelAttribution.pdf>.
- Ambler, T. (2000). Persuasion, pride and prejudice: how ads work, International Journal of Advertising, Vol. 19(3), pp. 299-315. <https://doi.org/10.1080/02650487.2000.11104803>.
- Anderl, E., 2014. Three essays on analyzing and managing online consumer behavior.
- Anderl, E., Becker, I., Von Wangenheim, F. & Schumann, J.H. (2016). Mapping the customer journey: Lessons learned from graph-based online attribution modeling, International Journal of Research in Marketing, Vol. 33(3), pp. 457-474.
- Anderl, E., Becker, I., Wangenheim, F.V. & Schumann, J.H. (2014). Mapping the Customer Journey: A Graph-Based Framework for Online Attribution Modeling, Rochester, NY, Available: <https://papers.ssrn.com/abstract=2343077>.
- Archak, N., Mirrokni, V. & Muthukrishnan, S. (Apr 26, 2010). Mining advertiser-specific user behavior using adfactors, Proceedings of the 19th international conference on world wide web, ACM, pp. 31-40.
- Arrate, A., Cabañas, J.G., Cuevas, Á, Calderón, M. & Cuevas, R. (2018). Large-scale analysis of user exposure to online advertising in Facebook, IEEE Access, Vol. 7 pp. 11959-11971. <https://arxiv.org/abs/1811.10921>.
- Bukiet, B., Harold, E.R. and Palacios, J.L., 1997. A Markov chain approach to baseball. Operations Research, 45(1), pp.14-23.
- Chandler-Pepelnjak, J. (2009). Measuring ROI beyond the last ad, Oxford University Press, US, What everyone needs to know Available: [https://ebookcentral.proquest.com/lib/\[SITE_ID\]/detail.action?docID=4883792](https://ebookcentral.proquest.com/lib/[SITE_ID]/detail.action?docID=4883792).
- Chatterjee, P., Hoffman, D.L. & Novak, T.P. (2003). Modeling the Clickstream: Implications for Web-Based Advertising Efforts, Marketing Science, Vol. 22(4), pp. 520-541. <https://pubsonline.informs.org/doi/10.1287/mksc.22.4.520.24906>.
- Dalessandro, B., Perlich, C., Stitelman, O. & Provost, F. (2012). Causally motivated attribution for online advertising, Proceedings of the Sixth International Workshop on data mining for online advertising and internet economy, ACM, pp. 1-9.
- Facebook (2019). About ad auctions, <https://www.facebook.com/business/help/430291176997542>.

Fawcett, T. (2006). An introduction to ROC analysis, Vol. 27(8)(Pattern recognition letters), pp. 86-874.

Google Ads Help About data-driven attribution

Google Analytics Help (2019). Cross Device reports with data from users who turned on Ads Personalization, <https://support.google.com/analytics/answer/7668466>.

Google Attribution Playbook (2012). Google Attribution Playbook, Index Press, Auckland,

Google Developers: Cookies and User Identification (2018). in: Google Developers,

Google (2019). About Google Analytics, <https://marketingplatform.google.com/about/analytics/>.

Data-driven attribution methodology (2019).

Jansen, B.J. & Schuster, S. (2011). Bidding of the buying funnel for sponsored search and keyword advertising, Journal of electronic commerce research, Vol. 12(1), pp. 1-18. <http://www.econis.eu/PPNSET?PPN=668112492>.

Jayawardane, C.H.W., Halgamuge, S.K. & Kayande, U. (2015). Attributing Conversion Credit in an Online Environment: An Analysis and Classification, 2015 3rd International Symposium on Computational and Business Intelligence (ISCBI), IEEE, pp. 68-73.

Kireyev, P., Pauwels, K. & Gupta, S. (2013). Do display ads influence search? <http://www.econis.eu/PPNSET?PPN=737694033>.

Noble, S. (2010). It's Time To Bury The Marketing Funnel, <http://www.iimagineservicedesign.com/wp-content/uploads/2015/09/Customer-Life-Cycle-Journey.pdf>,

European Digital Advertising market has doubled in size in 5 years | IAB Europe (2018).

IAB Internet Advertising Revenue Report (2018). in: Professional Services Close - Up, Close-Up Media, Inc, Jacksonville,

R Foundation The R Project for Statistical Computing <https://www.r-project.org/>.

Rentola, O. (2014). Analyses of Online Advertising Performance Using Attribution Modeling,

Sebastian Cano-Berlangaa, Jose-Manuel Gimenez-Gomez & Cori Vilella (2017). Attribution models and the Cooperative Game Theory,

Shao, X. & Li, L. (2011). Data-driven multi-touch attribution models, ACM, pp. 258-264.

Shapley, L.S. (1953). A value for n-person games, Contributions to the Theory of Games, Vol. 2(28), pp. 307-317.

Spedicato, G.A., Kang, T.S., Yalamanchi, S.B., Yadav, D. and Cordón, I., (2016). The markovchain package: a package for easily handling Discrete Markov Chains in R.

https://cran.rstudio.org/web/packages/markovchain/vignettes/an_introduction_to_markovchain_package.pdf

The Interactive Advertising Bureau (2016). Digital attribution primer 2.0, Available: <https://www.iab.com/wp-content/uploads/2016/10/Digital-Attribution-Primer-2-0-FINAL.pdf>.

Wilbur, K.C. & Zhu, Y. (2009). Click Fraud, Marketing Science, Vol. 28(2), pp. 293-308. <http://mktsci.journal.informs.org/cgi/content/abstract/28/2/293>.

Zhao, K., Mahboobi, S.H. & Bagheri, S.R. (2018). Shapley Value Methods for Attribution Modeling in Online Advertising, <https://www.openaire.eu/search/publication?articleId=od18::821c2344d96e4122c3b46f814eaefb54>

